AD-A115 562   AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOO--ETC F/G 12/1
COMPARISON OF ESTIMATION TECHNIQUES FOR THE FOUR PARAMETER BETA--ETC(U)
DEC 81 D E BERTRAND
UNCLASSIFIED   AFIT/GOR/MA/81D-1   NL

1 of 1

END
DATE FILMED
7 82
DTIC

AD A115562

COMPARISON OF ESTIMATION
TECHNIQUES FOR THE
FOUR PARAMETER BETA DISTRIBUTION

THESIS

AFIT/GOR/MA/81D-1

David E. Bertrand
2Lt USAF

AFIT/GOR/MA/81D-1

COMPARISON OF ESTIMATION TECHNIQUES FOR THE FOUR PARAMETER BETA DISTRIBUTION

THESIS

Presented to the Faculty of the School of Engineering

of the Air Force Institute of Technology

Air University

in Partial Fulfillment of the

Requirement for the Degree of

Master of Science

by

David E. Bertrand, B.S.

2LT USAF

Graduate Operations Research

December 1981

Preface

My interest in statistics in general began with courses taught here at AFIT by Lt. Col. Jim Bexfield, Dr. Albert Moore, and Capt. Brian Woodruff. I am grateful to them for teaching me the foundations on which this thesis was built. Dr. Moore, my thesis advisor, is due individual thanks for his guidance and support. I would like to thank Dr. J. P Cain, my reader, for his help in this effort. The aid and advice of James Sweeder, a PhD student of Dr. Moore, was invaluable and is greatly appreciated. I wish to take this opportunity to thank all of the AFIT Faculty, and my GOR-81D classmates, for their help and friendship over the past eighteen months.

Finally, I would like to thank my fiancee, Sharon Sorano, for enduring a long-distance relationship for the past year and a half. The support and encouragement that she has given me has been helpful beyond words.

David E. Bertrand

Contents

List of Figures

List of Tables

Abstract

This thesis compares three estimation techniques in application to the beta distribution: method of moments, maximum likelihood, and minimum distance. The four parameter version of the beta distribution is used; it has two shape parameters, and upper and lower limit parameters. Linear interpolation on order statistics is used to find initial estimates of the limits. The classical estimation procedures, method of moments and maximum likelihood, are applied through procedures found in the literature. A newer technique, minimum distance, is applied for the first time to the beta distribution.

Comparision of estimation techniques is accomplished using Monte Carlo analysis. Five sample sizes are considered -- 4, 8, 12, 16, and 20 -- and three pairs of shape parameters -- (3,3), (9,4), and (1,2) -- for a total of fifteen cases. One thousand samples are generated for each case, and each estimation technique is then applied to all samples. Two effectiveness measures are used; they are the mean square error of each parameter estimate, and the Cramer-von Mises distance between the estimated and the true distribution. These effectiveness measures are compared in each case to determine which technique provides the best overall effectiveness.

COMPARISON OF ESTIMATION TECHNIQUES
FOR THE FOUR PARAMETER BETA DISTRIBUTION

I. Introduction

Statistical estimation is currently used in private industry, in government, and in the military. Areas of application include quality control, logistics, and simulation. As estimation theory has been studied over time, different techniques have been developed for finding estimates of the parameters of probability distributions based on a sample from that distribution. Therefore, there is a need to perform a comparison of estimation methods for specific distributions and determine whether any one method out-performs the others. The research performed for this thesis undertakes such a comparison for the four-parameter beta distribution; the estimation methods compared are the method of moments, maximum likelihood, and minimum distance.

The following hypothetical situation illustrates how the results of this thesis might be used. A large international conglomerate, known as the Bertrand Corporation, produces a highly complex, technologically sophisticated piece of equipment called the widget. The president of the corporation, a very wise and knowledgeable man, realizes that his customers will require information on how long their widgets will last. He therefore would like to know the probability distribution of the time to failure (TTF) of his product.

The desired information could be obtained in a number of ways. First, every widget produced could be operated until it failed, and the length of operation recorded. This technique would provide perfect information on the TTF of each item; however, Bertrand Corporation stock could be expected to drop drastically due to lack of sales.

A less costly method would be to start by assuming that the TTF is normally distributed. Then, a random sample of the widgets could be taken from those produced, and run to failure. The mean and variance of this sample could then be used as estimates of the mean and variance of the underlying normal distribution. The difficulty inherent in this method is that the normality assumption may not be valid. If the analyst has no idea of the shape of the underlying distribution, he or she cannot be sure whether or not a normal curve can be made to fit it with reasonable accuracy.

The third method is to assume as the underlying distribution one which can take on a large variety of shapes, and then take a random sample of widgets from which to find estimates of the parameters of this distribution. The president of Bertrand Corporation knows that the beta distribution can take on many shapes, but also realizes that there are several methods available to perform the estimation. He therefore would like to know what method will give him the most accurate information on the time to failure of the widgets, so that he can pass this information on to his customers.

This thesis is undertaken in order to provide the aforementioned president, or anyone in a similar situation, the answer that he or she requires. The two classical methods of estimation, method of moments and maximum likelihood, and a more recent technique called minimum distance estimation, will be used to estimate the parameters of the four parameter beta distribution. The techniques will then be compared with each other, to determine if any one method provides superior results. This thesis report will proceed in four parts. First, the three estimation techniques will be reviewed in general. Second, the beta distribution will be described and application of the estimation methods to it will be discussed. Third will be a summary of the Monte Carlo analysis performed in order to evaluate the techniques and make the desired comparisons. Last, the results and conclusions of this thesis will be presented along with suggestions for future work on the subject.

II. Estimation Techniques

In the introduction, it was stated that this thesis would compare method of moments, maximum likelihood, and minimum distance estimation of the beta distribution. This chapter will provide some background and general theory on these three estimation techniques. First, however, a few words about estimation in general are in order. Estimation involves finding approximations, or estimates, of the parameters of a probabililty distribution through the use of estimators. Mendenhall and Scheaffer define an estimator as ''a rule that tells us how to calculate an estimate based on the measurements contained in a sample'' (Ref 19:264). The estimation process, then, is one of taking a sample from the population of interest, performing calculations on the sample points according to the ''rule'' of the estimator, and using the results of these calculations as the estimators of the parameters of the underlying distribution.

Next, this chapter will consider in detail three specific rules, or techniques, used for estimation. The first two are known as the classical estimation techniques; the third is a more recent method.

Method of Moments

''The method of moments is one of the oldest estimation techniques'' (Ref 4:7). The basic idea is fairly simple; use as estimators those values of the parameters for which sample moments equal population moments (Ref 19:300). The kth sample moment is computed from the sample as follows (Ref 19:300):

$$m'_k = \frac{1}{n}\sum_{i=1}^{N} X_i^k \qquad (2.1)$$

where n is the sample size and X_i is the ith sample point. The kth population moment is derived from the underlying distribution using the formula (Ref 19:300):

$$\mu'_k = E(X^k) \qquad (2.2)$$

The population moment will therefore be a function of the parameters of the underlying distribution. Let p be the number of parameters to be estimated. Setting up the equations

$$\mu'_k = m'_k \qquad k=1,\ldots,p \qquad (2.3)$$

provides a system of p equations in the p unknown parameters, which may be solved to find the method of moments estimators of the parameters.

It is sometimes convenient to use functions of the moments, rather than the moments themselves, when performing method of moments estimation. This can be done, provided the total number of different moments used is still equal to the number of parameters being estimated. Two such functions which are commonly used are the skewness and the kurtosis. Before defining these, it is necessary to define a different class of moments, known as the central moments. The kth central moment, or moment about the mean, of the sample is (Ref 18:18):

$$m_k = \frac{1}{n}\sum_{i=1}^{N}(X_i - \bar{X})^k \tag{2.4}$$

where $\bar{X}$ is the sample mean ($\bar{X} = m_i' = m_i$). The kth population central moment is (Ref 18:18):

$$\mu_k = E((X - \mu)^k) \tag{2.5}$$

where $\mu = E(X)$ is the mean of the distribution. The central moments are used in finding the sample skewness, $\hat{K}_3$, and population skewness, K_3, as follows (Ref 18:20-21):

$$\hat{K}_3 = m_3/(m_2)^{3/2} \tag{2.6}$$

$$K_3 = \mu_3/(\mu_2)^{3/2} \tag{2.7}$$

The sample and population kurtoses are derived from the following formulas, the '^' again indicating the sample statistic (Ref 18:20-21):

$$\hat{K}_4 = m_4/(m2)^2 \tag{2.8}$$

$$K_4 = \mu_4/(\mu_2)^2 \tag{2.9}$$

Since the first and second moments, skewness, and kurtosis together involve only the first four moments, they can be used to find moment estimates for four parameters.

Maximum Likelihood

The method of maximum likelihood uses as estimators the parameter values which maximize the likelihood, or joint density, of the sample (Ref 19:303). Let $f(y;\underline{\theta})$ be the underlying probability distribution, where $\underline{\theta} = (\theta_1,\theta_2,\ldots,\theta_p)$ is a vector of parameter values. Since the sample is chosen at random, the joint density of the sample is merely the product of the probability distribution evaluated at each of the sample points (Ref 19:171); the likelihood is therefore defined by the formula,

$$L(X_1,\ldots,X_n;\theta) = \prod_{i=1}^{n} f(X_i;\theta) \qquad (2.10)$$

The goal is to find the vector $\hat{\theta}$ which maximizes this function. This can be done by setting each of the p partial derivatives $\partial L/\partial\theta_k$ $k=1,\ldots,p$ to zero and solving for the $\hat{\theta}_k$. In practice, it is usually expedient to take the natural logarithm of L before maximizing; this transforms the product into a sum, which is easier to differentiate. Maximizing ln(L) will result in the same values for $\hat{\theta}_k$, since the natural logarithm is a monotonically increasing function (Ref 19:303). Therefore, the maximum likelihood estimators are the roots of the p simultaneous equations

$$\frac{\partial}{\partial \theta_k} \ln L(x_1, \ldots, x_n; \theta_1, \ldots, \theta_p) = 0 \quad , k = 1, \ldots, p \qquad (2.11)$$

In some cases, the estimators cannot be solved for in closed form and must be found by iteration (Ref 4:6).

Minimum Distance

Minimum distance estimation has been developed more recently than the previous two methods; its development took place in the 1950's. This method evolved from attempts by estimation theorists to strike a balance between the characteristics of robustness and consistency. The term robustness refers to the ability of an estimator to adapt to deviations in the underlying model and remain efficient (Ref 21:3). An estimator is consistent if it converges in probability to the true value of the parameter as the sample size tends to infinity (Ref 19:309). The initial work in minimum distance estimation was performed by J. Wolfowitz. He published a paper in 1953 (Ref 26), and another in 1957 (Ref 27) which outlined the minimum distance method and showed it to be consistent; "in a wide variety of cases, [the minimum distance method] will furnish super-consistent estimators even when classical methods...fail to give consistent estimators" (Ref 26:9).

It has not been until recently, however, that minimum distance estimation has begun to be widely applied. In their 1979 paper, Parr and Schucany applied the method to estimation of the location parameter of a symmetric distribution, emphasizing the normal distribution, and found it to yield ''*strongly consistent estimators with excellent robustness properties*'' (Ref 21:5). Other applications have been accomplished at the Air Force Institute of Techology (AFIT) under the supervision of Dr. Albert H. Moore. They are estimation of the location parameters of the generalized exponential power distribution by Maj. Larry McNeese (Ref 17), estimation of the parameters of the generalized t distribution, by Capt. Tony Daniels (Ref 4), estimation of the three parameter weibull distribution, by Capt. Robert Miller (Ref 20), and estimation of the three parameter gamma distribution, by Capt. William L. James (Ref 13). These studies generally found the minimum distance estimators to be better than the classical methods. This thesis is a continuation of these efforts.

The minimum distance method is an extention of the goodness of fit tests used in testing hypotheses. To test, by goodness of fit, the hypothesis that a sample is from a certain distribution with certain parameter values, one constructs the distribution function $F(x;\underline{\Theta})$ at these parameter values, and determines how well it fits the distribution function of the sample (called the empirical distribution function, or EDF) by some previously defined measure of fit. Commonly, the goodness of fit measure is some measure of the distance between $F(x;\underline{\Theta})$ and the EDF. Minimum distance estimation merely takes as its estimate of $\underline{\Theta}$ those values which minimize the distance between $F(x;\underline{\theta})$ and $Sn(x)$.

Minimum distance estimation requires that three things be specified: the family of distribution functions $F(X;\underline{\theta})$, a rule for obtaining the empirical distribution function, denoted $Sn(x)$, and a measure of the distance between $F(x;\underline{\theta})$ and $Sn(x)$. Previous applications to other families of distribution have already been mentioned; this thesis deals with the beta family of distributions. There are several EDF's which can be used, among them the 1/n step function and median ranks. The distance measure is of crucial importance; often, distance estimators are identified by the name of the distance measure used. They will be described in more detail.

The most common distance measures are descibed in Parr and Schucany; the following definitions are from this paper (Ref 21:7-8). The first is the weighted Kolmogorov distance

$$D_{\xi}(Sn,F) = \sup_{x\varepsilon R}\left|Sn(x)-F(x;\theta)\right|\xi(F(x;\theta)) \qquad (2.12)$$

where 'sup' signifies the least upper bound and ξ is a weighting function. This statistic should be familiar to those experienced with the Kolmogorov-Smirnov goodness of fit test (Ref 3:347). Another measure developed from a goodness of fit statistic is the weighted Cramer-von Mises distance

$$W_{\xi}^{2}(Sn,F) = \int_{-\infty}^{\infty}(Sn(x)-F(x,\underline{\theta}))^{2}\xi(F(x,\underline{\theta}))dF(x,\underline{\theta}) \qquad (2.13)$$

where, again, ξ is a weighting function. Use of uniform weighting, $\xi(\cdot)=1$, defines the unweighted Cramer-von Mises (CVM) measure $W^2(Sn,F)$. The weighting scheme $\xi(F) = \frac{1}{F(1-F)}$, $0<F<1$ defines the Anderson-Darling distance measure, which is denoted $A^2(Sn,F)$. Kuiper's maximal interval probability distance is given by the equation,

$$V(Sn,F) = \sup_{-\infty<a<b<\infty} \left| (Sn(b)-Sn(a)) - (F(b;\underline{\theta})-F(a;\underline{\theta})) \right| \quad (2.14)$$

Last, a general class of distance measures is defined by

$$Z_{a,b}(Sn,F) = a\int_{-\infty}^{\infty}(Sn(X)-F(X;\theta))^2 dF(X;\theta)+b\left[\int_{-\infty}^{\infty}(Sn(X)-F(X;\theta))dF(X;\theta)\right]^2 \quad (2.15)$$

This class includes the CVM measure when a=0 and b=1, Watson's measure, denoted $U^2(Sn,F)$ when a=1 and b=1, and Chapman's measure, when a=0 and b=1. The previous AFIT studies which were mentioned have used the Kolmogrov, Cramer-von Mises, and Anderson-Darling distance measures.

III. The Beta Distribution

> The beta distribution is becoming more and more widely used in applied statistical analysis in many business disciplines. For example, in finance the beta distribution has been employed in an attempt to measure the probability of payment or default in a credit granting decision. In management, the beta distribution is often used in PERT. And in marketing, the beta distribution is frequently employed in Markovian brand-switching models when transition probabilities are taken to be random variables rather than parameters (Ref 8:1).

These are just a few ways in which the beta distribution is applied in the "real world." However, work on estimation of the four parameter beta, with all four parameters unknown, is scarce; in preparation for this thesis, only one paper--by Glenn E. Tarr (Ref 24)--was found which dealt with estimation of all four parameters. Virtually all of the work done on estimation of the beta that this author encountered either dealt only with the two parameter version or assumed that the other two parameters were known constants.

This chapter deals with the four parameter beta distribution: what it is, and how the three estimation techniques described in Chapter II may be applied to it. The first section of this chapter defines the beta family of probability distributions and reviews some of its characteristics. The second section descibes the method used to get preliminary estimates for two of the parameters, which are required in applying the three estimation procedures in Chapter II. The final three sections of this chapter consider how the method of moments, maximum likelihood, and minimum distance are applied to the beta distribution.

General

This thesis applies the three estimation techniques explained in the previous chapter to the beta family of probability distributions. The most general form of the probability density function (pdf) is referred to as the four parameter beta distribution, and has the form (Ref 14:37):

$$f(y;P,Q,A,B) = \begin{cases} \frac{1}{B(P,Q)} (Y-A)^{P-1}(B-Y)^{Q-1}/(B-A)^{P+Q-1}, & A \leq Y \leq B \\ 0 \text{ , otherwise} \end{cases} \qquad (3.1)$$

The parameters A and B define the range over which the function is defined; hence, any random variate from this distribution must fall between A and B. The parameters P and Q are known as the shape parameters, since they determine the shape of the graph of the probability density function. Both P and Q are required to be strictly greater than zero, and B must be strictly greater than A. Exchanging the values of P and Q cause the graph to be reflected about the midpoint of the line segment AB. The function B(P,Q) is defined by the formula (Ref 19:130):

$$B(P,Q) = \int_0^1 Y^{P-1}(1-Y)^{Q-1}dY = \frac{\Gamma(P)\Gamma(Q)}{\Gamma(P+Q)} \qquad (3.2)$$

The special case where A=0 and B=1 is commonly known as the standard or two parameter beta, and has the form (Ref 14:37):

$$f(y;P,Q,) = \begin{cases} \frac{1}{B(P,Q)} X^{P-1}(1-X)^{Q-1}, & 0 \leq X \leq 1 \\ 0, & \text{otherwise} \end{cases} \quad (3.3)$$

If A and B are known, a random variable from the four parameter beta can be transformed into a standard beta random variable by the equation (Ref 14:37):

$$X = \frac{Y-A}{B-A} \quad (3.4)$$

Some members of the beta family have specific names attached to them. When Q=1 in equation 3.1, it is sometimes called the power function, or in equation 3.3, the standardized power function (Ref 14:37). The standard beta with P=Q=1/2 is known as the arc-sine distribution, and is used in random walk theory *(Ref 14:39)*. When P=Q=1, the beta distibution reduces to the well-known continuous uniform, or rectangular, distribution.

The cumulative distribution function (cdf) of the beta distribution is found by integrating equation 3.1 from A to the point at which the cdf is to be evaluated. The cdf of the standard beta is commonly called the incomplete beta function, and is denoted $I_y(P,Q)$ (Ref 19:131).

As mentioned in the introduction, the main reason that the beta family of distributions is useful in fitting empirical distribution functions is its ability to take on many different shapes. A few of these shapes are presented in Figure 3-1. A more complete collection of graphs of beta density functions is included in Johnson and Koltz (Ref 14:42-44). When both P and Q are less than one, the function is U-shaped. When one is less than and the other greater than one, it is J-shaped. If both P and Q are greater than one, the function is bell-shaped. If P is greater than Q, the function is skewed to the right, the opposite if Q>P. The function is symmetric about (B-A)/2 if P equals Q.

Obtaining Starting Values

The solution methods used to accomplish the three estimation techniques explained in Chapter II all require an initial estimate of the parameters A and B. This is done through interpolation performed on the sample points as follows. First, the sample is sorted from smallest to lagest, so that $X_{(i)}$ is the ith order statistic (Ref 19:229). Then the median rank is found for the two smallest and two largest points. The median rank of $X_{(i)}$ is computed using the formula (Ref 15:31):

$$MR(X_{(i)}) = \frac{i-0.3}{n+0.4} \qquad (3.5)$$

For convenience, $MR(X_{(i)})$ shall be denoted Y_i.

The interpolation method is displayed graphically in

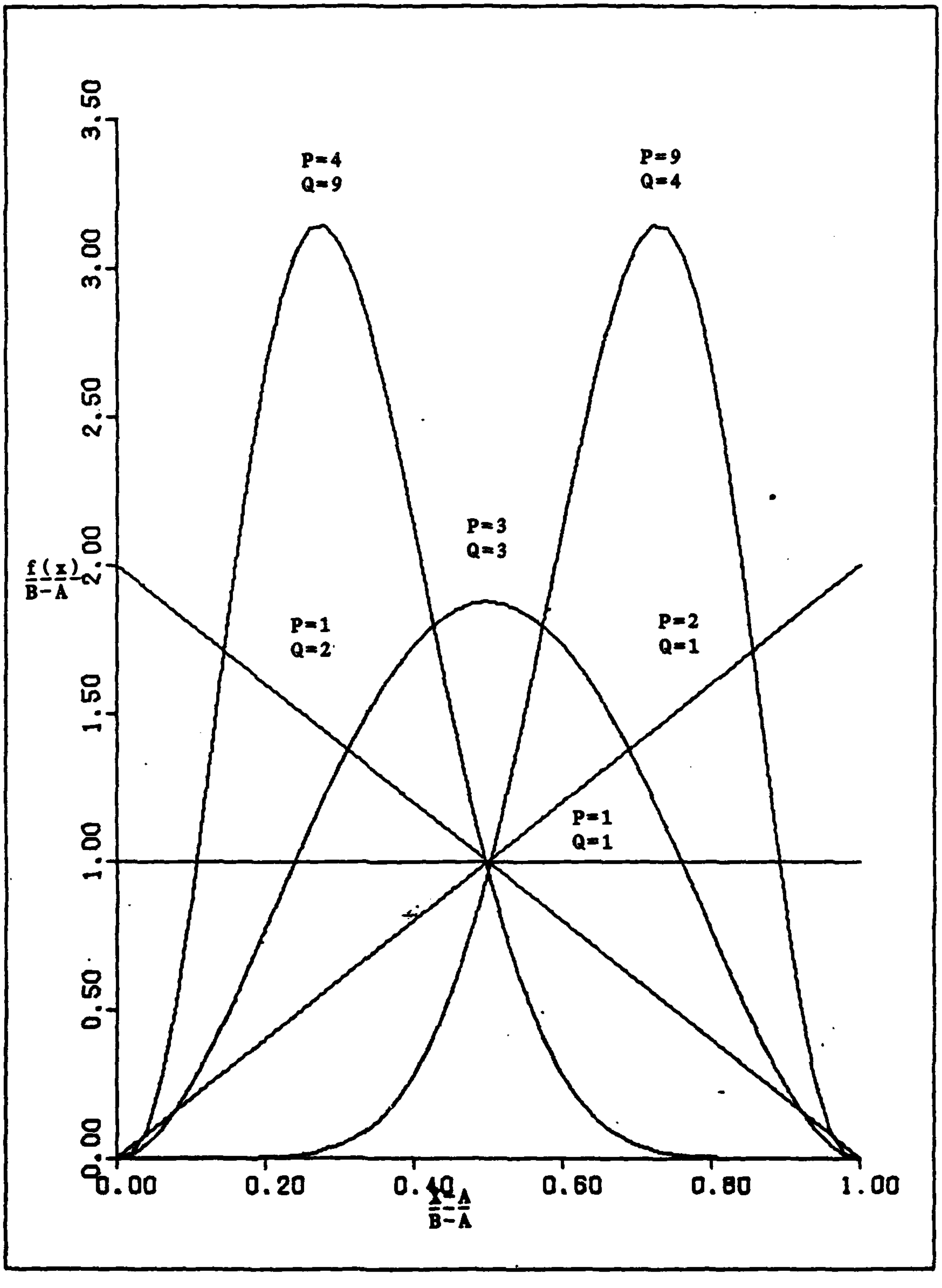

Figure 1. Some Shapes of the Beta Distribution

Figure 3-2. In finding the estimate for A, the slope of the line connecting the first two points is calculated by the usual formula:

$$m = (Y_2 - Y_1)/(X_{(2)} - X_{(1)}) \tag{3.6}$$

The estimate for A is the point at which this line intercepts the x-axis. Using this slope formula on this lower portion of the line, and then solving for A, provides the following:

$$\hat{A} = X_{(1)} - Y_1/M \tag{3.7}$$

The same procedure performed on the largest two order statistics gives the formula for estimating B:

$$\hat{B} = (1 - Y_n)/m + X_{(n)} \tag{3.8}$$

where m is calculated using the nth and n-1st points, in place of the second and first in equation 3.6.

To the author's knowledge, this is a new method for finding these parameters of the beta distribution. It was not mentioned in any of the literature that was read in preparation for this thesis. It should be noted that interpolation will always give plausible estimates for A and B; that is, $\hat{A}$ is always smaller than the first order statistic, and $\hat{B}$ is always larger than the last. The estimates would be

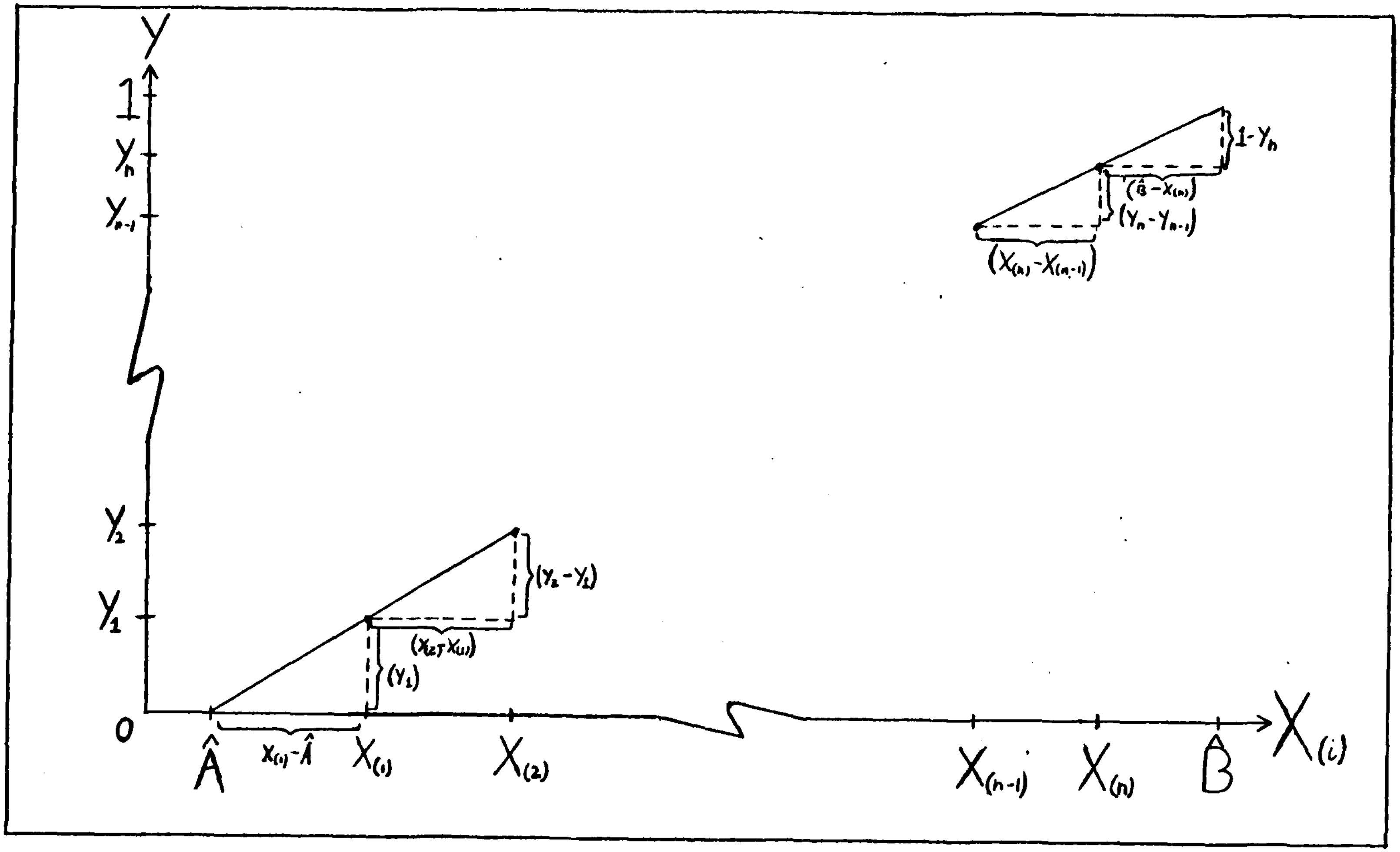

Figure 2. Interpolation for Starting Values

expected to be more accurate with larger sample sizes. Consistency, although not proven here, seems apparent; as n increases, the first two order statistics would draw closer to the true value of A, and therefore the li_ear approximation would better fit the tail of the true cdf. An equivalent argument applies to B on the upper side. Intuitively, it seems reasonable to expect the estimate of A to be more accurate that $\hat{B}$ when the pdf is skewed left -- that is, when Q>P -- and to expect $\hat{B}$ to be more accurate than $\hat{A}$ when the pdf is skewed right -- when P>Q. This would occur because the skewness causes a large portion of the data to be at one end of the range, providing a closer interpolation at that end.

<u>Method of Moments</u>

The method of moments is the only estimation technique for which an application to the full four parameter beta was found in the literature; that being a paper by Glenn Tarr (Ref 24). Moment estimation of the standard beta is discussed in detail by Fielitz and Myers (Ref 8), and the formulas for the two moments required to estimate the standard beta are available from this and various other sources (Ref 14,18).

There are two possible approaches to performing moment estimation of the beta distribution. First, the estimates of A and B found by interpolation could be kept as final estimates, and the first two moments used to find P and Q in a manner similar to Fielitz and Myers. The second method is to use the first four moments to acquire method of moments estimators for all four of the parameters. Both of these alternatives will be evaluated in this thesis.

The first method requires the formulas for the first two moments of the beta distribution. The central moments will be used; the formulas are as follows (Ref 14:44):

$$\mu_1 = E(X) = A + \frac{P(B-A)}{(P+Q)} \qquad (3.9)$$

$$\mu_2 = Var(X) = [(B-A)^2 PQ]/[(P+Q)\ (P+Q+1)] \qquad (3.10)$$

These formulas are then equated to the first sample moment, the sample mean X, and the second sample central moment, the (biased) sample variance s^2. These equations are then solved for P and Q, resulting in the folowing equations:

$$P = [(\bar{X}^*)^2(1-\bar{X}^*)]/S^{2*} - \bar{X}^* \qquad (3.11)$$

$$Q = [(1-\bar{X}^*)^2\bar{X}^*]/S^{2*} - (1-X^*) \qquad (3.12)$$

where $\bar{X}^* = (\bar{X}-A)/(B-A)$ is the standardized sample mean, and $S^{2*} = S^2/(B-A)^2$ is the standardized sample deviation. Using the interpolated estimates of A and B, the values of P and Q are thus easily found from the data. In Tarr's paper, he suggests using the two moment method for ''relatively large'' samples, and suggests the first and last order statistics as estimates for A and B. (Ref 24:3).

The above technique is only a ''partial'' method of moments estimation, since moments are not used to estimate A and B. Glenn Tarr, in the paper referred to earlier, estimated all four parameters by method of moments, using the skewness and kurtosis. The population skewness and kurtosis for the beta distibution are given by the formulas (Ref 18:44):

$$K_3 = [2(P-Q)(P+Q+1)^{1/2}]/[(PQ)^{1/2}(P+Q+2)] \quad (3.13)$$

$$K_4 = 3(P+Q+1)[2(P+Q)^2+PQ(P+Q-6)]/PQ(P+Q+2)(P+Q+3) \quad (3.14)$$

In Tarr's paper, he used a shifted kurtosis; his population kurtosis formula is equivalent to subtracting three from equation 3.14. His formula for the sample skewness and kurtosis are as follows (Ref 24:8):

$$\hat{K}_3 = [n^3T_3 - 3nT_2T_1 + 2T_1^3]/SD^3n(n-1)(n-2) \quad (3.15)$$

$$\hat{K}_4 = [(n^3+n^2)T_4 - 4(n^2+n)T_3T_1 - 3(n^2-n)T_2^2 + 12nT_2T_1^2 - 6T_1^4]/ \\ [SD^4n(n-1)(n-2)(n-3)] \quad (3.16)$$

Where $T_y = \sum_{i=1}^{n} X_i^q$, and SD is the sample standard deviation.

Since the skewness and kurtosis involve only P and Q, setting these formulas equal to the sample skewness and kurtosis respectively provides two equations in two unknowns which may be solved by numerical techniques. Once these values are found, A and B are estimated using equations 8 and 9; solving for A and B rather than P and Q leads to the formulas:

$$\hat{A} = \bar{X} - ([S^2\hat{P}(\hat{P}+\hat{Q}+1)]/\hat{Q})^{1/2} \qquad (3.17)$$

$$\hat{B} = [\bar{X}(\hat{P}+\hat{Q})-\hat{A}\hat{Q}]/\hat{P} \qquad (3.18)$$

Thus, the method of moment estimators have been formed for all four parameters.

<u>Maximum</u> <u>Likelihood</u>

To the author's knowledge, maximum likelihood estimation of the complete four parameter beta distribution has not yet been developed. No reference to an ML technique which estimates A and B were found in the literature. This thesis will deal with a "partial" maximum likelihood estimation process; that is, one which keeps the interpolated estimates of A and B as constants and estimates only P and Q by maximum likelihood.

The technique used for maximum likelihood estimation of the beta is that developed by Gnanadesikan, Pinkham, and Hughes in 1967 (Ref 11). It dealt with the standard beta, and performed the estimation using smallest order statistics.

The resulting simultaneous equations were solved by Newton's method, and it was stated that "The starting values...used are crucial for the efficient convergence of the iterative scheme" (Ref 11:611).

Beckman and Tietjen picked up on Gnanadesikan, et al., and developed a solution method which is "fast, simple," and for which "No starting values are required and no convergence problems have been encountered" (Ref 2:254). The likelihood equations to be solved are given as follows (Ref 2:253):

$$\xi(\hat{P}) - \xi(\hat{P}+\hat{Q}) = \ln G_1 \qquad (3.19)$$

$$\xi(\hat{Q}) - \xi(\hat{P}+\hat{Q}) = \ln G_2 \qquad (3.20)$$

where

$$G_1 = \prod_{i=1}^{n} [(X_i-A)/(B-A)]^{1/n} \qquad (3.21)$$

$$G_1 = \prod_{i=1}^{n} [(B-X_i)/(B-A)]^{1/n} \qquad (3.22)$$

$$\xi(Z) = \frac{d}{dZ}\ln(\Gamma(Z)) \quad \text{(Ref 11:609)} \qquad (3.23)$$

In order to solve the simultaneous equations 3.15 and 3.16, Beckman and Tietjen used the following procedure. First, equation 3.16 was solved for $\xi(\hat{P}+\hat{Q})$. This was substituted into 3.15, which was then solved for $\xi(\hat{P})$. Then, the inverse of $\xi(\cdot)$ was taken of each side, providing an equation for $\hat{P}$. This equation was then substituted for $\hat{P}$ in equation 3.16, leading to the following equation (Ref 2:254):

$$\xi(\hat{Q})-\xi\{[\xi^{-1}(\ln G_1-\ln G_2+\xi(\hat{Q}))]+\hat{Q}\}-\ln G_2 = 0 \quad (3.24)$$

The root of this equation was found by the secant method; this same method was used to evaluate $\xi^{-1}(\cdot)$. The function $\xi(\cdot)$ was evaluated using an approximation given in the reference. The secant method "requires the user to specify an interval...within which the root is located" (Ref 2:254). Beckman and Tietjen provide tables from which $\hat{P}$ and $\hat{Q}$ can be found for given values of G_1 and G_2; a listing of their computer program is also provided. This thesis uses that program to find the maximum likelihood estimates of $\hat{P}$ and $\hat{Q}$, given the interpolated estimates of A and B.

In a comment on the Fielitz and Myers paper (Ref 8), which favored the method of moments for estimating the beta distribution, and on a rebuttal by Romesburg (Ref 22), which supported maximum likelihood, Kottas and Lau (Ref 16) wrote an excellent article which summarizes both classical methods of estimating the beta, provides historical perspective, and comments on which is the better technique for estimating P and Q. They state that Fisher, the initial developer of the maximum likelihood method, "mathematically proved that the inherent variance of an MM estimator [of the beta distribution] is always greater than or equal to that of the corresponding ML estimator and approaches the latter only in near normal cases'' (Ref 16:529). In an article which focused on the small sample case, Dishon and Weiss reached a similar conclusion; they compiled a table comparing the MM and ML estimators for various parameter values and sample sizes, and concluded that ''with few exceptions the ML estimator is more accurate for low n than is the moment estimator'' (Ref 5:4). In order for the results of this thesis to be consistent with these findings, the ''partial'' maximum likelihood estimates of P and Q would be expected to be more accurate than the ''partial'' moment estimates.

Minimum Distance

This thesis represents, to the author's knowledge, the first attempt at estimation of the parameters of the beta distribution by the minimum distance technique. It is the first attempt at AFIT to estimate more than one parameter by minimum distance. Although some of the AFIT theses, mentioned earlier (Refs 4, 13, 17, 20), did deal with more than one parameter, only one parameter was estimated by minimum distance; this estimate was then used to improve the estimates of the other parameters found by other methods. This thesis will attempt to find minimum distance estimates for all four parameters of the beta distribution.

The empirical distribution function to be used in this thesis is the 1/n step function, which assigns the ith point of the ordered sample the value i/n. The Cramer-von Mises distance measure will be used. When applied to the 1/n step function, the CVM measure defined in equation 2.13, with uniform weighting, reduces to (Ref 23:731):

$$W^2(Sn,F) = \sum_{i=1}^{N}\left[F(X_i) - \frac{2i-1}{2n}\right]^2 \qquad (3.25)$$

Where $F(X_i)$ is the cdf evaluated at the ith sample point. This same source provides similar formulas for applying many of the other distance formulas mentioned in Chapter II to the 1/n step function EDF.

The process used to find the minimum distance estimates is as follows. First, the initial estimates of A and B are found by interpolation. Then the moment equations, 3.11 and 3.12, are used to get starting values for P and Q. Holding A and B fixed, equation 3.25 is minimized for P and Q (the parameters P, Q, A, and B are implicitly contained in $F(X_i)$). After this minimization is accomplished, P and Q are held constant at the new values, and 3.25 is minimized for A and B. The resulting values of P, Q, A, and B are the minimum distance estimates of these parameters.

IV: Monte Carlo Analysis

This chapter deals with the specific method used to perform the comparison of estimation techniques which is the main purpose of this thesis. The comparison is performed using Monte Carlo analysis. There are basically three steps to a Monte Carlo anaysis of an estimation method. First, random samples from the distribution to be estimated are generated. Second, the parameters of the distribution are estimated from these samples. Third, the estimations are evaluated as to how well they approximated the true distribution. This chapter will discuss each of these steps in detail.

Since there are a vast amount of data and large numbers of calculations involved in Monte Carlo analysis, use of a high-speed computer is a necessity. A Control Data Corporation (CDC) computer system, located at Aeronautical Systems Division, Wright-Patterson Air Force Base, Ohio, was used in performing the analysis for this thesis. In programming each of the three steps outlined above, existing software was used whenever possible; specifically, subroutines from the International Mathematical Statistics Library (IMSL) were widely used. The reader should refer to the IMSL manual (Ref 12) if specific information about these routines is desired.

Generation of Data

In order for the comparisons made by this thesis to be valid, they should be made for a number of sample sizes and several different combinations of parameter values. Five sample sizes were used: 4, 8, 12, 16, and 20. For each of these sample sizes, three combinations of P and Q were used: P=3, Q=3; P=9, Q=4; and P=1, Q=2 (refer to Figure 1 in Chapter III for a graph of these curves). In order to save on computer time, only one combination of values was used for A and B; different values are not expected to have an affect on estimation ability, since such a change would only result in a linear translation along the axis. The values A=2, B=10 were arbitrarily chosen for this analysis.

For each of the 15 cases (five sample sizes times three P, Q combinations), 1000 samples were generated for use in estimation. Generation of beta random variates was accomplished using the IMSL routine GGBTR, which provides an array of standard beta random variates for a specified P, Q, and sample size. The resulting array was then sorted using the IMSL routine VSRTA. The random variates were then unstandardized, using equation 3.4 when solved for Y instead of X, so that the random variates now formed a sample from the desired four parameter beta distribution.

Since the interpolated estimates of A and B are needed for all three estimation techniques, they were calculated in the same program which performed the data generation. The method is exactly as described in Chapter III; the median ranks of the first two and last two points of the previously sorted sample were calculated using equation 3.5, the slopes were found using equation 3.6 and a similar expression for the two highest points, and then equations 3.7 and 3.8 were applied to calculate the estimates. The mean and standard deviation of each sample were also calculated at this time. The number of replications (1000 in this analysis), sample size, true values of P, Q, A, and B, and 1000 sets of replication number, list of random variates, interpolated estimates of A and B, mean, and standard deviation were stored on permanent file for use by each of the three estimation programs. A listing of the program which performed this data generation is provided in Appendix C.

Computerization of Estimation Techniques

Method of Moments. In the previous chapter, it was explained that two sets of moment estimators would be calculated. They are the 'partial' MM estimators, using the interpolated estimates of A and B and finding $\hat{P}$ and $\hat{Q}$ by the first two moments, and the 'full' MM estimators, which use the first two moments, skewness, and kurtosis to calculate $\hat{P}$, $\hat{Q}$, $\hat{A}$ and $\hat{B}$. The computer program written to do this is included in Appendix D.

The 'partial' method of moments estimates were found first; this was done through direct application of equations 3.11 and 3.12. The 'full' MM estimation technique is based on the procedure suggested by Tarr (Ref 24). Finding $\hat{P}$ and $\hat{Q}$ from the skewness and kurtosis involves solving a system of two nonlinear equations in two unknowns. The IMSL routine ZSCNT was used to accomplish this, using the 'partial' MM estimates of P and Q as starting values. If the program were to attempt to estimate either P or Q as less than zero, the program was designed to use the 'partial' moment estimators for that sample. When the program was executed, this was found to occur for virtually every sample; when attempts to remedy this failed, Tarr's approach was abandoned. These results will be discussed further in the next chapter.

Maximum Likelihood. The program used to perform the ML estimation is taken from the article by Beckman and Tietjen (Ref 2:258). The only changes made were for input of data, calculation of G_1 and G_2, and output of results. The reader should refer to the source article for a program listing; since the changes were superficial, the listing will not be provided here.

Minimum Distance. Computerization of the minimum distance method follows directly from the process outlined in Chapter III. After the data was read in, the IMSL routine ZXMIN was used to perform the minimization. This routine minimizes a function, in this case equation 3.25, for an array of parameters. Parr and Schucuny used this routine in their analysis (Ref 21:21). Using the two-moment estimates as starting values, $\hat{P}$ and $\hat{Q}$ were used first as input parameters for ZXMIN, while $\hat{A}$ and $\hat{B}$ were held constant through use of a COMMON statement. When this minimization was completed, ZXMIN was used a second time, this time with $\hat{A}$ and $\hat{B}$

as input parameters, using the interpolated estimates as starting values, while the values of $\hat{P}$ and $\hat{Q}$ found in the first minimization were held constant using a COMMON statement. The estimate of A was set equal to the first order statistic if ZXMIN attempted to estimate A greater than $X_{(1)}$; similarly, $X_{(n)}$ was used for $\hat{B}$ if ZXMIN attempted to set $\hat{B}$ < $X_{(n)}$. The listing of the computer program used to perform the minimum distance estimation is included in Appendix E.

Comparison of Estimation Techniques

The third step in the Monte Carlo analysis is to evaluate the estimates; this evaluation will be used as a basis in comparing the estimation techniques. There are two approaches which could be used for this evaluation. The first approach would be to individually measure how close the estimates of each parameter are to the true value of that parameter. The measure commonly used for this type of evaluation is the mean square error (MSE). The second approach is to calculate an overall measure of how well the estimated distribution fits the true distribution. A distance measure of the type outlined in Chapter III is an appropriate measure for this approach.

This thesis used both of these approaches; the mean square errors of $\hat{P}$, $\hat{Q}$, $\hat{A}$ and $\hat{B}$ were found, and the mean CVM distance between the estimated and true cdf calculated, for each estimation method and each of the 15 cases of sample size and P, Q values. The program written to perform both of these evaluations is listed in Appendix F. These two approaches will now be explained in more detail.

Mean Square Errors. The mean square error is a measure, based on repeated estimation, of how well an estimation method has estimated a given parameter. The formula for calculating the mean square error is as follows:

$$\mathrm{MSE}(\hat{\theta}) = \left[\sum_{i=1}^{N}(\hat{\theta}_i-\theta)^2\right]/N \qquad (4.1)$$

where θ is the true value of the parameter, $\hat{\theta}_i$ is the ith estimate, and N is the number of times the estimation is performed--in this analysis, N=1000. A difficulty in using MSE's when many parameters are estimated is that conflicting results are possible; estimation method A may have the smallest MSE for parameter 1, while method B has the smallest MSE for parameter 2 for the same case. Another potential difficulty with MSE's is that they are not scale invariant; the same size MSE may be highly significant for a small valued parameter, but insignificant for a larger one. This can complicate comparison of MSE's for different parameter values.

CVM Distance. The Cramer-von Mises (CVM) goodness of fit statistic W^2 was defined by equation 2.13, using the uniform weighting $\xi(\cdot)=1$. In this usage, however, the estimated cdf $\hat{F}$ is used in place of Sn; the integral is multiplied by the sample size to form the actual distance measure (Ref 23). Since $dF(x)=\frac{dF}{dx}dx=f(x)dx$, the formula for the distance between the estimated cdf $\hat{F}$ and the true cdf F is:

$$W^2(\hat{F},F) = n\int_A^B(\hat{F}(X,\hat{\underline{\theta}})-F(X,\underline{\theta}))^2 f(X,\underline{\theta})dX \qquad (4.2)$$

where $\underline{\theta}=(P,Q,A,B)$, $\hat{\underline{\theta}}=(\hat{P},\hat{Q},\hat{A},\hat{B})$, and $f(x,\theta)$ is the true probability density function (pdf). This integral was evaluated using 16 point Gaussian quadrature. This solution technique requires that the integral be over the limits -1 to 1; this is satisfied through the identity (Ref 10:221):

$$\int_a^b g(x)dx = \frac{b-a}{2}\int_{-1}^{1} g\left[\frac{(b-a)t+b+a}{2}\right]dt \qquad (4.3)$$

where

$$X = \frac{(b-a)t+b+a}{2}$$

Gaussian quadrature estimates an integral with limits -1 and 1 using a weighted sum, as follows (Ref 1:916):

$$\int_{-1}^{1} g(x)dx \sim \sum_{i=1}^{n} w_i g(x_i) \qquad (4.4)$$

where w_i and x_i are the ith Gaussian weights and quadrature points respectively, and n is the number of points; for this analysis, n=16. The appropriate weights and points were taken from the Handbook of Mathematical Functions (Ref 1:916). Since it is possible that the estimates of A and B will be inside of the true values, some of the quadrature points may be in places where $\hat{F}$ is not defined. This problem was overcome by defining $\hat{F}(x)$ to be zero when $X<\hat{A}$ and one when $X>\hat{B}$. This situation is depicted graphically in Figure 3. Having the estimates outside of the true values of A and B presents no problem, since the integral is calculated only between the true values.

The CVM statistic was calculated as just described for each of the 1000 replications. The sample mean and standard deviation of the CVM statistics can therefore be calculated using the usual formulas:

$$\bar{W}^2 = \left[\sum_{i=1}^{1000} W_i^2\right]/1000 \tag{4.5}$$

$$SD(W^2) = \sum_{i=1}^{1000}(W_i^2-\bar{W}^2)^2/1000 \tag{4.6}$$

where W_i^2 is the CVM distance of the ith estimation.

Since the number of replications is large, the central limit theorem implies that the mean CVM statisic is appoximately normally distributed (Ref 19:252). Therefore, a confidence interval for the mean CVM statistic may be calcu-

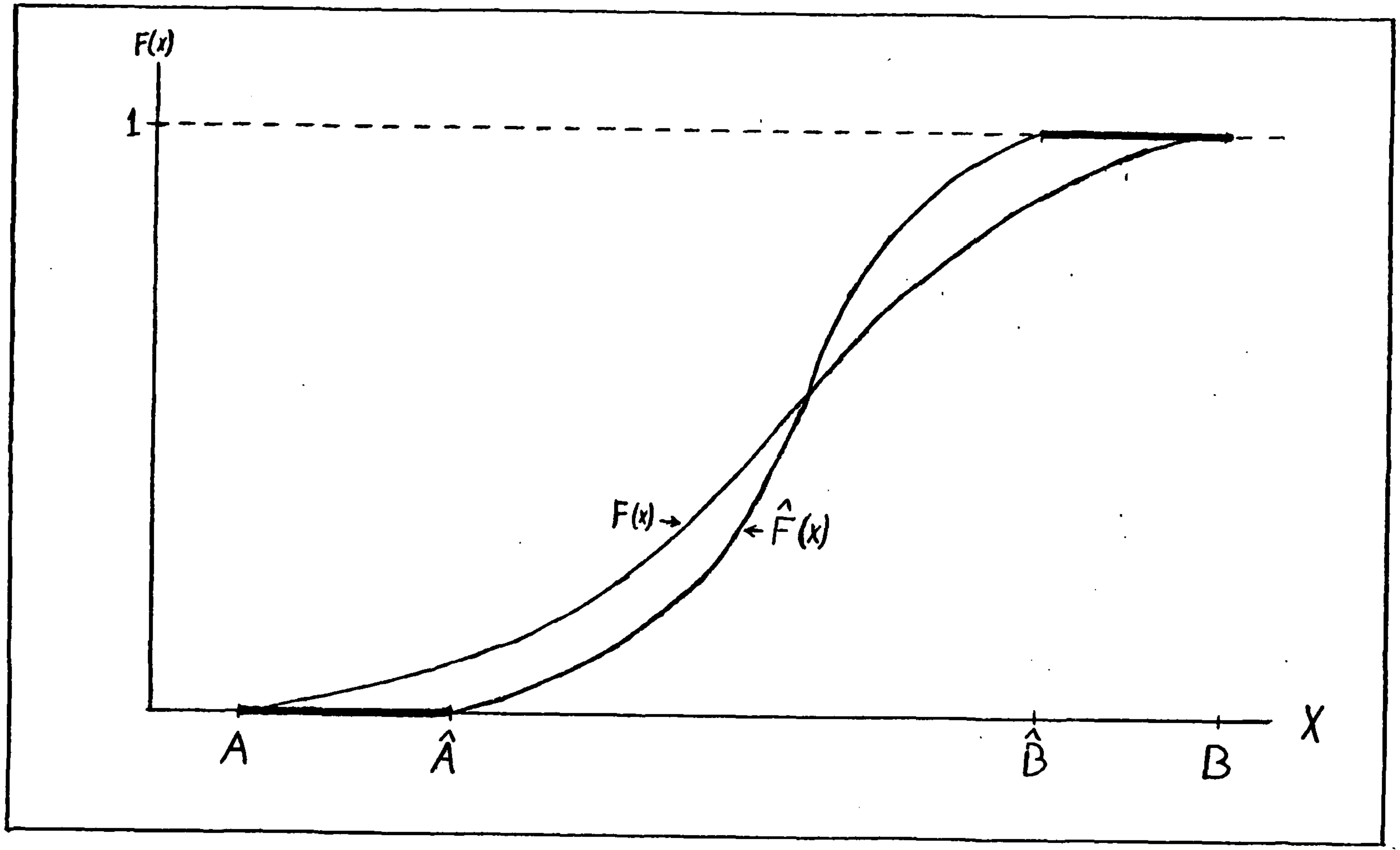

Figure 3. Estimated and True CDF Curves

lated (Ref 19:277). The formula for the confidence interval is as follows (Ref 25:195):

$$P(\bar{W}^2 - Z_{\alpha/2} SD(W^2)/\sqrt{N} < E(W^2) < \bar{W}^2 + Z_{\alpha/2} SD(W^2)/\sqrt{N}) = 1-\alpha \quad (4.7)$$

where α is the significance level, $Z\alpha/2$ is the value of the standard normal leaving an area of $\alpha/2$ to the right, and N is the number of replications, in this case, 1000.

The *CVM distance* has *the advantage over* the MSE of being a single measure of fit for all four parameters, and also of being scale invariant with respect to the size of the parameter values. One disadvantage, which goes along with the *first* advantage, is that information about the individual parameters is lost. At times, one may wish to know which method worked better on a particular parameter; for this, the MSE is the better measure. It should be clear that *there* is no bias introduced by using the CVM measure both in finding the minimum distance estimate and then again in evaluating this estimate. That is because during minimization the distance between the estimated distribution function and the *empirical* distribution *function* is *measured*, while during evaluation the distance measured is between the estimated distribution function and the true distribution function.

V. Results and Conclusions

Results of Comparisons

The numerical results of the two comparison approaches are summarized in the appendices. There are fifteen tables in each appendix, one for each case of sample size and parameter values. Appendix A contains the tables of mean square errors. The three estimation methods; method of moment estimation (MME), maximum likelihood estimation (MLE), and minimum distance estimation (MDE) are listed down the side. The mean square errors of $\hat{P}$ (MSEP), $\hat{Q}$ (MSEQ), $\hat{A}$ (MSEA), and $\hat{B}$ (MSEB) are then listed across the row for each method. Appendix B contains the means, standard deviations, and confidence intervals for the CVM statistics. Again, the three methods are listed down the side. Across each row are, in order, the sample mean of the CVM statistics (MCVM), the sample standard deviation of the CVM statistics (SDCVM), the lower limit and the upper limit of the 95% confidence interval of the mean of the CVM distance (95% C.I.). It should be noted that SDCVM is an estimate of the population standard deviation; it must be divided by $\sqrt{1000}$ to get an estimate of the standard deviation of the mean.

The reader may note that, in some cases, the MSE's and the mean CVM statistic seem to contradict each other; that is, one method may have MSE's equal to or smaller than another for all parameters, but the other method has a smaller mean CVM statistic. The case of N=4, P=9, Q=4 is such a case; both MME and MLE are the interpolated estimates of A and B, so these MSE's are equal. However, the MSE'S of both $\hat{P}$ and $\hat{Q}$ are smaller for the MLE, but the mean CVM is smaller for the MME.

This phenomenon can occur because of the way in which the four parameters of the beta distribution interact. The reader should refer to Figure 1 in Chapter III, and review the shape of the beta when P=9 and Q=4. The curve remains very close to the x-axis until about 1/4 of the way from A to B. For this reason, A is usually interpolated at about this point on the axis. Using this as the lower limit, however, the distribution is much more symmetric than when the true value of A is used. Therefore, the values of $\hat{P}$ and $\hat{Q}$ which best approximate the curve for the interpolated estimates of A and B are different from the true values of P and Q. A set of values which are further away, in an MSE sense, from the true values than another set, may thus be closer in a distance sense to the true distribution due to the estimation of A and B.

Since the MSE's and the mean CVM statistics contradict each other at times, it is necessary to choose one to serve as a basis of comparison between the estimation methods. The mean CVM statistic will be used since it is an overall measure and does not depend on the particular parameter values. The method with the smallest mean CVM statistic provides the closest fit, on the average, to the true distribution. Using this criterion, the moment estimate is ranked first, followed by maximum likelihood and then minimum distance, for all sample sizes in cases P=Q=3 and P=9, Q=4; and when N=4 for P=1, Q=2. For the other four sample sizes with P=1, Q=2, maximum likelihood has the smallest mean CVM, with MME next, and then MDE. THe MSE, however, is smaller for MLE than for MME in most of the cases.

Using just the point estimates of the mean CVM distance gives no indication of whether the differences between the MCVM for each method is large enough to be significant. For this, the 95% confidence intervals can be used. If the confidence intervals of two methods overlap, this indicates that the mean CVM distances of the methods are not significantly different. This comparison is equivalent to performing a t test of the difference between the means. Since there are three means being considered, comparing these three confidence intervals equates to performing multiple t tests; for this reason, the effective α level - that is, the probability of finding two means to be significantly different when they are not - is actually somewhat higher than 0.05. When the confidence intervals listed in Appendix B are compared, they show that the 95% confidence intervals for the MME and MLE distance measures overlap in every case. This indicates that the difference between them is not statistically significant. Comparing confidence intervals for the MDE indicates mixed results. In some cases (N=4, P=3, Q=3; N=4, P=1, Q=2; N=12, P=9, Q=4), the CVM statistic of the minimum distance method is significantly greater than for both of the other methods. In five cases (N=4, P=9,

Q=4; N=8, P=9, Q=4; N=12, P=3, Q=3; N=12, P=1, Q=2; N=20, P=3, Q=3), the CVM statistic for MDE is significantly greater than the smaller of the other distances, but not the larger one. In the other seven cases, all three confidence intervals overlap, so that none of the differences are significant.

Although not used in comparing overall effectiveness, the mean square errors must be used if effectiveness in estimating a particular parameter is to be compared. Problems arise, however, in comparing the MSE's of $\hat{P}$ and $\hat{Q}$, since they depend on the values of $\hat{A}$ and $\hat{B}$ as described earlier. However, since the interpolated estimates of A and B do not depend on other parameters, these can be compared to the MDE estimates of A and B using MSE's. In this comparison, the MDE fares better than in the previous paragraph. Overall, the differences in the MSE's are small. The interpolated estimates are better in all cases with sample size four. For sample size eight, neither method has a clear superiority. With N=12, the MSE's using minimum distance are smaller in all cases except for $\hat{A}$ when P=1, Q=2, where it is slightly larger. The minimum distance estimates have a smaller MSE than the interpolated estimates of A and B for all cases with sample sizes 16 and 20.

In Chapter III, in the section on obtaining starting values, a number of suppositions were made concerning the interpolated values of $\hat{A}$ and $\hat{B}$. The results in Appendix A can now be used to test these suppositions. Consistency was the first supposition; the accuracy of the estimates was expected to increase as sample size increasd. This is supported by the results. The mean square errors of both $\hat{A}$ and $\hat{B}$ decrease monotonically as the sample size increases for all three sets of parameter values. The second supposition was that left-skewed distributions would provide better estimates of A, while right-skewed distributions would provide better estimates of B. This is also borne out by the results. For the case P=9, Q=4, which is skewed right, the MSE of $\hat{B}$ is always much less than the MSE for $\hat{A}$; in fact, the MSE of $\hat{A}$ is always at least nine times the MSE of $\hat{B}$. For the left-skewed case, P=1, Q=2, the MSE of $\hat{A}$ is always less than the MSE of $\hat{B}$ by at least a factor of seven. In the symmetric case, P=Q=3, the MSE's of $\hat{A}$ and $\hat{B}$ are nearly equal in all the sample sizes.

Conclusions

Does the president of the Bertrand Corporation have an answer to his question? What is the best method? Based upon the results of this analysis, and for the range of sample sizes considered, the method of moments using interpolated values of A and B seems to the the best choice of the three methods investigated. It provides estimates which fit the true distribution at least as well as the other two methods, and is more easily computed than either the methods of maximum likelihood or minimum distance. For a single sample, the moment estimates may be easily found using a desk calculator. The method of maximum likelihood provides estimates of nearly equal accuracy to the method of moments, and is certainly a viable alternative; however, the computation is more involved and, for an exact solution, requires computerization. The minimum distance method, when applied to all four parameters, does not appear to be as good, especially in light of the fact that it requires much more computer time than the others to obtain a solution. The fact that its estimates of A and B were improvements would seem to indicate that minimum distance might be successfully applied to the location-type parameters of the beta distribution.

The four moment estimation technique suggested by Tarr (Ref 24) is apparently not as straightforward as he seems to believe. Careful re-reading of this paper revealed that the author apparently did no actual verification or Monte-Carlo analysis of this technique at all. His tables seem to have been generated merely by choosing values of P and Q, plugging them into the formulas for the population skewness and kurtosis, and tabling the resulting values. More work is required on this method if it is to be a viable alternative to the others presented herein.

Recommendations for Further Study

As stated, investigation into the viability of Tarr's method is a possible area of study. Another possibiliy would be extension of the maximum likelihood method in Beckman and Tietjen (Ref 2) to all four parameters. A third alternative would be to combine MME and MLE into a 'hybrid' approach. This method could use the interpolated values of $\hat{A}$ and $\hat{B}$ to obtain ML estimates of $\hat{P}$ and $\hat{Q}$, just as done in this thesis. Then, these estimates of $\hat{P}$ and $\hat{Q}$ could be used in equations 3.17 and 3.18 to obtain new estimates of A and B with the first two moments. It would then be possible to use these to re-estimate P and Q, and perhaps loop through this procedure until the desired accuracy is achieved. The study would have to determine if this would lead to significant improvements over the methods evaluated in this thesis.

Regarding minimum distance estimation, this was the first attempt at applying this method to the beta distribution, and it should not be abandoned just because it is not yet as good as the other techniques. There is still work to be done. One area that could be explored would be to try other distance measures to see if they may lead to an improvement. For instance, the Anderson-Darling statistic may turn out to be better for estimating A and B, since it is more heavily weighted at the tails of the distribution. Another possibility would be to only estimate one parameter by MDE. One may wish to use Tarr's formulation for the beta distribution, which uses X_o for A as the location parameter, and M, which equals B-A, as a range parameter (Ref 24:1). X_o and M could be found by interpolation, $\hat{P}$ and $\hat{Q}$ by MM or ML, and then X_o could be refined by MD, keeping $\hat{M}$, $\hat{P}$, and $\hat{Q}$ fixed. This new $\hat{X}_o$ could then be used to improve the estimates of $\hat{P}$ and $\hat{Q}$. This would also cut down on the use of computer time, since only one parameter, rather than four, would be estimated by minimum distance.

Bibliography

1. Abramowitz, M. and Irene A. Stegun. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. New York: Dover Publication, Inc., 1964.

2. Beckman, R. J. and G. L. Tietjen. "Maximum Likelihood Estimation for the Beta Distribution," Journal of Statistical Computation and Simulation, 7: 253-258 (1978).

3. Conover, W. J. Practical Nonparametric Statistics (Second Edition). New York: John Wiley & Sons Inc., 1980.

4. Daniels, Tony G., Capt. Robust Estimation of the Generalized t Distribution Using Minimum Distance Estimation. *MS Thesis*. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December 1980.

5. Dishon, M. and G. H. Weiss. "Small Sample Comparison of Estimation Methods for the Beta Distribution," Journal of Statistical Computation and Simulation, 11: 1-11 (1980).

6. Easterling, Robert G. "Goodness of Fit and Parameter Estimation," Technometrics, 18: 1-9 (February 1976).

7. Elderton, William P. Frequency Curves and Correlation. Washington, D. C.: Harren Press, 1953.

8. Fielitz, B. D. and B. L. Meyers. "Estimation of Parameters in the Beta Distribution," Decision Sciences, 6 (1): 1-13 (January 1975).

9. -----. "Estimation of Parameters in the Beta Distribution-Reply," Decision Sciences, 7 (1): 163-164 (January 1976).

10. Gerald, C. F. Applied Numerical Analysis (Second Edition). Reading, Massachusetts: Addison-Wesley Publishing Company, 1978.

11. Gnanadesikan, R., R. S. Pinkham and L. P. Hughes. "Maximum Likelihood Estimation of the Parameters of the Beta Distribution from the Smallest Order Statistics," Technometrics, 9: 607-620 (1967).

12. IMSL Library Reference Manual. Houston: IMSL, Inc., 1980.

13. James, William L., Capt. Robust Minimum Distance Estimation Based on a Family of Three-Parameter Gamma Distributions. MS Thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December 1980.

14. Johnston, N. L. and Samuel Kotz. Continuous Univariate Distributions, Volume 2. Boston: Houghton Mifflin Co., 1970.

15. Kaper, K. C. and L. R. Lamberson. Reliability in Engineering Design. New York: John Wiley & Sons, Inc., 1977.

16. Kottas, J. F. and H. S. Lau. ''On Estimating Parameters for Beta Distibutions,'' Decision Sciences, 9 (3): 526-531 (July 1978).

17. McNeese, Larry B., Major. Adaptive Minimum Distance Estimation Techniques Based on a Family of Generalized Exponential Power Distributions. MS Thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December 1980.

18. Merkle, Robert G. Statistical Measures, Probability Densities, and Mathematical Models for Stochastic Measurements. AFFDL-TR-76-83. Wright-Patterson AFB, Ohio: Air Force Flight Dynamics Laboratory, October 1976.

19. Mendenhall, William and Richard L. Scheaffer. Mathematical Statistics with Applications. North Scituate, Massachusetts: Duxbury Press, 1973.

20. Miller, Robert M., Capt. Robust Minimum Distance Estimation of the Three Parameter Weibull Distribution. MS Thesis. Wright-Patterson AFB, Ohio: Air Force Institute of Technology, December 1980.

21. Parr, William C., and William R. Schucany. Minimum Distance and Robust Estimation. Southern Methodist University, Dallas, Texas: Department of Statistics, 1979.

22. Romesburg, H. C. ''Estimation of Parameters in the Beta Distribution-Comment,'' Decision Sciences, 7 (1): 162 (January 1979).

23. Stephens, M. A. ''EDF Statistics for Goodness of Fit and Some Comparisons,'' Journal of the American Statistical Association, 69: 730-737 (September 1974).

24. Tarr, Glenn E. ''The Beta Distribution: Estimating Four Parameters by the Method of Moments,'' unpublished article. Springfield, Illinois: Sangamo Electric Company, date unknown.

25. Walpole, Ronald E. and R. H. Myers. Probability and Statistics for Engineers and Scientists (Second Edition). New York: MacMillan Publishing Co., Inc., 1978.

26. Wolfowitz, J. ''Estimation by the Minimum Distance Method,'' Annals of Mathematical Statistics, 5: 9-23 (1953).

27. -----. ''The Minimum Distance Method,'' Annals of Mathematical Statistics, 28: 75-88 (1957).

Appendix A

Tables of Mean Square Errors

The following notation is used in this Appendix.

Term	Notation
Method of Moments Estimation	MME
Maximum Likelihood Estimation	MLE
Minimum Distance Estimation	MDE
Mean Square Error of $\hat{P}$	MSEP
Mean Square Error of $\hat{Q}$	MSEQ
Mean Square Error of $\hat{A}$	MSEA
Mean Square Error of $\hat{B}$	MSEB

Monte Carlo sample size is 1000 and true values of A and B are 2 and 10 for all tables.

TABLE A-I

Mean Square Errors

Sample size	4
True P	3
True Q	3

	MSEP	MSEQ	MSEA	MSEB
MME	3.742108	3.658086	4.668322	4.653648
MLE	2.739536	2.649655	4.668322	4.653648
MDE	12.371356	14.542017	4.744504	4.737503

TABLE A-II

Mean Square Errors

Sample Size	4
True P	9
True Q	4

	MSEP	MSEQ	MSEA	MSEB
MME	60.861580	8.601137	16.911807	1.726146
MLE	55.594636	6.942450	16.911807	1.726146
MDE	67.128382	13.105965	17.014121	1.756985

TABLE A-III

Mean Square Errors

Sample size	4
True P	1
True Q	2

	MSEP	MSEQ	MSEA	MSEB
MME	0.410244	0.936122	1.217973	8.993774
MLE	0.626805	0.645073	1.217973	8.993774
MDE	5.416876	21.291772	1.239676	9.114334

TABLE A-IV

Mean Square Errors

Sample size 8
True P 3
True Q 3

	MSEP	MSEQ	MSEA	MSEB
MME	3.427817	3.354894	2.904387	2.940100
MLE	3.049777	2.987382	2.904387	2.940100
MDE	9.766766	9.877664	2.890916	2.957770

TABLE A-V

Mean Square Errors

Sample Size 8
True P 9
True Q 4

	MSEP	MSEQ	MSEA	MSEB
MME	56.805536	7.960173	13.934737	1.117937
MLE	54.788050	7.319176	13.934737	1.117937
MDE	60.792229	18.236968	13.786411	1.111636

TABLE A-VI

Mean Square Errors

Sample size 8
True P 1
True Q 2

	MSEP	MSEQ	MSEA	MSEB
MME	0.332060	0.975033	0.335525	4.998726
MLE	0.328806	0.862249	0.335525	4.998726
MDE	8.262157	23.410740	0.363392	4.971074

TABLE A-VII

Mean Square Errors

Sample size 12
True P 3
True Q 3

	MSEP	MSEQ	MSEA	MSEB
MME	2.825820	2.868660	2.089426	2.126352
MLE	2.761430	2.815917	2.089426	2.126352
MDE	5.539327	4.950296	2.013997	2.032997

TABLE A-VIII

Mean Square Errors

Sample Size 12
True P 9
True Q 4

	MSEP	MSEQ	MSEA	MSEB
MME	52.775368	7.208131	12.263634	0.858765
MLE	52.519962	7.059538	12.263634	0.858765
MDE	49.354920	9.095091	12.013108	0.826734

TABLE A-IX

Mean Square Errors

Sample size 12
True P 1
True Q 2

	MSEP	MSEQ	MSEA	MSEB
MME	0.230257	0.838969	0.142688	3.206688
MLE	0.207590	0.777774	0.142688	3.206688
MDE	1.013119	3.495416	0.146702	3.129456

TABLE A-X

Mean Square Errors

Sample size	16
True P	3
True Q	3

	MSEP	MSEQ	MSEA	MSEB
MME	2.609627	2.558815	1.826545	1.650983
MLE	2.708616	2.671452	1.826545	1.650983
MDE	2.943447	2.845556	1.732211	1.560583

TABLE A-XI

Mean Square Errors

Sample Size	16
True P	9
True Q	4

	MSEP	MSEQ	MSEA	MSEB
MME	50.354216	6.726745	11.302901	0.716087
MLE	51.150863	6.818277	11.302901	0.716087
MDE	46.387784	6.755438	11.102680	0.690490

TABLE A-XII

Mean Square Errors

Sample size	16
True P	1
True Q	2

	MSEP	MSEQ	MSEA	MSEB
MME	0.168482	0.782000	0.093014	2.697157
MLE	0.155296	0.777950	0.093014	2.697157
MDE	0.364832	1.755449	0.092245	2.604381

TABLE A-XIII

Mean Square Errors

Sample size	20
True P	3
True Q	3

	MSEP	MSEQ	MSEA	MSEB
MME	2.300955	2.353331	1.454751	1.414300
MLE	2.483768	2.551249	1.454751	1.414300
MDE	2.645247	2.566947	1.369513	1.325909

TABLE A-XIV

Mean Square Errors

Sample Size	20
True P	9
True Q	4

	MSEP	MSEQ	MSEA	MSEB
MME	49.079484	6.438000	10.798213	0.640201
MLE	50.543684	6.711302	10.798213	0.640201
MDE	43.965367	5.839172	10.579763	0.610275

TABLE A-XV

Mean Square Errors

Sample size	20
True P	1
True Q	2

	MSEP	MSEQ	MSEA	MSEB
MME	0.125813	0.645373	0.057157	2.090599
MLE	0.133994	0.639272	0.057157	2.090599
MDE	0.238399	1.041494	0.056218	2.006221

Appendix B

Tables of Cramer-von Mises Distance: Means, Standard Deviations, and Confidence Intervals

The following notation is used in this appendix.

Term	Notation
Method of Moments Estimation	MME
Maximum Likelihood Estimation	MLE
Minimum Distance Estimation	MDE
Mean CVM Distance	MCVM
Standard Deviation of CVM Distance ...	SDCVM
95% Confidence Interval upper and lower limits	95% C.I.

Monte Carlo sample size is 1000 and true values of A and B are 2 and 10 for all tables.

TABLE B-I

CVM Distance Statistics

Sample size	4
True P	3
True Q	3

	MCVM	SDCVM	95% C.I.
MME	0.135119	0.131025	0.126998--0.143240
MLE	0.139685	0.136396	0.131231--0.148139
MDE	0.161920	0.163691	0.151774--0.172066

TABLE B-II

CVM Distance Statistics

Sample size	4
True P	9
True Q	4

	MCVM	SDCVM	95% C.I.
MME	0.138114	0.132926	0.129875--0.146353
MLE	0.143925	0.140191	0.135236--0.152614
MDE	0.161689	0.162098	0.151642--0.171736

TABLE B-III

CVM Distance Statistics

Sample size	4
True P	1
True Q	2

	MCVM	SDCVM	95% C.I.
MME	0.140421	0.142939	0.131562--0.149280
MLE	0.143407	0.147824	0.134245--0.152569
MDE	0.164817	0.168862	0.154351--0.175283

TABLE B-IV

CVM Distance Statistics

Sample size	8
True P	3
True Q	3

	MCVM	SDCVM	95% C.I.
MME	0.128062	0.144414	0.119111--0.137023
MLE	0.129722	0.148687	0.120506--0.138938
MDE	0.144501	0.164251	0.134321--0.154681

TABLE B-V

CVM Distance Statistics

Sample size	8
True P	9
True Q	4

	MCVM	SDCVM	95% C.I.
MME	0.122436	0.128460	0.114474--0.130398
MLE	0.125101	0.132076	0.116915--0.133287
MDE	0.140132	0.146378	0.131059--0.149205

TABLE B-VI

CVM Distance Statistics

Sample size	8
True P	1
True Q	2

	MCVM	SDCVM	95% C.I.
MME	0.132165	0.140954	0.123429--0.140901
MLE	0.129635	0.144304	0.120691--0.138579
MDE	0.147750	0.164972	0.137525--0.157975

TABLE B-VII

CVM Distance Statistics

Sample size	12
True P	3
True Q	3

	MCVM	SDCVM	95% C.I.
MME	0.121909	0.138635	0.113316--0.130502
MLE	0.123787	0.140387	0.115086--0.132488
MDE	0.141451	0.167972	0.131040--0.151862

TABLE B-VIII

CVM Distance Statistics

Sample size	12
True P	9
True Q	4

	MCVM	SDCVM	95% C.I.
MME	0.116762	0.129083	0.108761--0.124763
MLE	0.118035	0.129308	0.110020--0.126050
MDE	0.138822	0.150955	0.129466--0.148178

TABLE B-IX

CVM Distance Statistics

Sample size	12
True P	1
True Q	2

	MCVM	SDCVM	95% C.I.
MME	0.127104	0.141719	0.118320--0.135888
MLE	0.119488	0.134865	0.111129--0.127847
MDE	0.141441	0.168347	0.131007--0.151875

TABLE B-X

CVM Distance Statistics

Sample size	16
True P	3
True Q	3

	MCVM	SDCVM	95% C.I.
MME	0.114084	0.132344	0.105881--0.122287
MLE	0.115781	0.129284	0.107768--0.123794
MDE	0.131064	0.154444	0.121491--0.140637

TABLE B-XI

CVM Distance Statistics

Sample size	16
True P	9
True Q	4

	MCVM	SDCVM	95% C.I.
MME	0.114048	0.132149	0.105857--0.122239
MLE	0.117757	0.135425	0.109363--0.126151
MDE	0.130373	0.149206	0.121125--0.139621

TABLE B-XII

CVM Distance Statistics

Sample size	16
True P	1
True Q	2

	MCVM	SDCVM	95% C.I.
MME	0.131006	0.149941	0.121713--0.140299
MLE	0.123323	0.143858	0.114407--0.132239
MDE	0.141448	0.168164	0.131025--0.151871

TABLE B-XIII

CVM Distance Statistics

Sample size	20
True P	3
True Q	3

	MCVM	SDCVM	95% C.I.
MME	0.109630	0.124946	0.101886--0.117374
MLE	0.112547	0.125415	0.104770--0.120316
MDE	0.128660	0.147926	0.119491--0.137829

TABLE B-XIV

CVM Distance Statistics

Sample size	20
True P	9
True Q	4

	MCVM	SDCVM	95% C.I.
MME	0.119805	0.135413	0.111412--0.128198
MLE	0.124453	0.137251	0.115946--0.132960
MDE	0.137788	0.154729	0.128198--0.147378

TABLE B-XV

CVM Distance Statistics

Sample size	20
True P	1
True Q	2

	MCVM	SDCVM	95% C.I.
MME	0.122389	0.135526	0.113989--0.130789
MLE	0.116783	0.130674	0.108684--0.124882
MDE	0.131730	0.148973	0.122497--0.140963

APPENDIX C

COMPUTER LISTING OF PROGRAM BETAGEN

```
      PROGRAM BETAGEN
C     ******************************************************************
C     *                                                                *
C     * WRITTEN BY 2LT DAVID E. BERTRAND AFIT/GOR-81D FOR MS THESIS    *
C     *                        DECEMBER 1981                           *
C     *                                                                *
C     * PURPOSE:    GENERATION OF SAMPLES OF BETA RANDOM VARIATES      *
C     *             CALCULATION OF MEAN AND STANDARD DEVIATION OF SAMPLES*
C     *             CALCULATION OF ESTIMATES OF A AND B BY INTERPOLATION *
C     *                                                                *
C     * VARIABLES: DSEED - SEED FOR RANDOM NUMBER GENERATOR            *
C     *            P     - FIRST SHAPE PARAMETER OF TRUE DISTRIBUTION  *
C     *            Q     - SECOND SHAPE PARAMETER OF TRUE DISTRIBUTION *
C     *            A     - LOWER LIMIT OF TRUE DISTRIBUTION            *
C     *            B     - UPPER LIMIT OF TRUE DISTRIBUTION            *
C     *            NR    - DESIRED SAMPLE SIZE                         *
C     *            NREPS - NUMBER OF SAMPLES TO BE GENERATED           *
C     *            GGBTR - IMSL ROUTINE WHICH GENERATES                *
C     *                    STD BETA VARIATES                           *
C     *            VSRTA - IMSL ROUTINE WHICH SORTS AN ARRAY           *
C     *                    INTO ASCENDING ORDER                        *
C     *            SUM   - DUMMY VARIABLE USED IN FINDING MEAN, STD DEV *
C     *            X     - ARRAY CONTAINING SAMPLE POINTS              *
C     *            MEAN  - ARITHMATIC MEAN OF SAMPLE                   *
C     *            SD    - STANDARD DEVIATION OF SAMPLE (BIASED)       *
C     *            Y1    - MEDIAN RANK OF FIRST ORDER STATISTIC        *
C     *            Y2    - MEDIAN RANK OF SECOND ORDER STATISTIC       *
C     *            YN1   - MEDIAN RANK OF N-1ST ORDER STATISTIC        *
C     *            YN    - MEDIAN RANK OF NTH ORDER STATISTIC          *
C     *            SLOPE - SEE EQUATION 3.6 IN THESIS                  *
C     *            ESTA  - INTERPOLATED ESTIMATE OF A: SEE EQN 3.7     *
C     *            ESTB  - INTERPOLATED ESTIMATE OF B: SEEE EQN 3.8    *
C     *                                                                *
C     * I/O FILES: INPUT - UNFORMATTED INPUT OF TRUE PARAMETER VALUES  *
C     *            TAPE5 - OUTPUT OF TRUE PARAMETERS, SAMPLES,         *
C     *                    CALCULATED VALUES                           *
C     *                                                                *
C     * IMPORTANT: IMSL LIBRARY MUST BE ATTACHED BEFORE PROGRAM IS RUN *
C     *            REVIEW IMSL MANUAL ON GGBTR AND VSRTA BEFORE RUNNING *
C     *                                                                *
C     ******************************************************************
```

```
      EXTERNAL  GGBTR,VSRTA
      DOUBLE PRECISION DSEED
      DIMENSION X(50)
      REAL  P,Q,MEAN
      INTEGER NR

      DSEED=1859217525.D0
C     READ PARAMETERS AND WRITE THEM TO FILE
      READ*,  P,Q,A,B,NR,NREPS
      WRITE(5,100) NREPS,NR,P,Q,A,B
  100 FORMAT(I4/I3/4(F10.6/))
C    ***** BEGIN LOOP FOR GENERATION OF SAMPLES *****
      DO 999 J=1,NREPS
       WRITE(5,103) J
  103  FORMAT(I4)
C      GENERATE AND SORT SAMPLE FROM STANDARD BETA
       CALL GGBTR(DSEED,P,Q,NR,X)
       CALL VSRTA(X,NR)
C      UNSTANDARDIZE RANDOM VARIATES,
C      WRITE DATA TO FILE
C      CALCULATE MEAN
       DO 10 I=1,NR
         X(I)=(B-A)*X(I)+A
         WRITE(5,101) X(I)
  101    FORMAT(F10.6)
         SUM=SUM+X(I)
   10  CONTINUE
       MEAN=SUM/NR
C      CALCULATE STANDARD DEVIATION
       SUM=0.0
       DO 20 I=1,NR
         SUM=SUM+(X(I)-MEAN)*(X(I)-MEAN)
   20  CONTINUE
       SD=(SUM/NR)**0.5
C      CALCULATE MEDIAN RANKS
C      INTERPOLATE TO ESTIMATE A AND B
       Y1=(1.-0.3)/(NR+0.4)
       Y2=(2.-0.3)/(NR+0.4)
       SLOPE=(Y2-Y1)/(X(2)-X(1))
C       EQUATION 3.7:
       ESTA=X(1)-Y1/SLOPE
       YN1=(NR-1-0.3)/(NR+0.4)
       YN=(NR-0.3)/(NR+0.4)
       SLOPE=(YN-YN1)/(X(NR)-X(NR-1))
C       EQUATION 3.8:
       ESTB=(1-YN)/SLOPE+X(NR)
C      WRITE CALCULATED VALUES TO FILE
       WRITE(5,102) ESTA,ESTB,MEAN,SD
  102  FORMAT(F10.6/F10.6/F10.6/F10.6)
C    ***** END LOOP *****
  999 CONTINUE
      STOP
      END
```

APPENDIX D

COMPUTER LISTING OF PROGRAM MOMENTS

PROGRAM MOMENTS

```
C     *******************************************************************
C     *                                                                 *
C     * WRITTEN BY 2LT DAVID E. BERTRAND AFIT/GOR-81D FOR MS THESIS     *
C     *                         DECEMBER 1981                           *
C     *                                                                 *
C     * PURPOSE:   TWO AND FOUR MOMENT ESTIMATION OF                    *
C     *            THE BETA DISTRIBUTION                                *
C     *                                                                 *
C     * VARIABLES: NREPS - # SAMPLES FOR WHICH ESTIMATION DONE (INPUT)  *
C     *            N     - SAMPLE SIZE (INPUT)                          *
C     *            PR    - TRUE VALUE OF FIRST SHAPE PARAMETER  (INPUT) *
C     *            QR    - TRUE VALUE OF SECOND SHAPE PARAMETER (INPUT) *
C     *            AR    - TRUE LOWER LIMIT OF DISTRIBUTION     (INPUT) *
C     *            BR    - TRUE UPPER LIMIT OF DISTRIBUTION     (INPUT) *
C     *            NDIV  - NUMBER OF TIMES 4 MOMENT ESTIMATION FAILS    *
C     *            K     - SAMPLE INDEX (INPUT)                         *
C     *            X     - ARRAY OF SAMPLE POINTS (INPUT)               *
C     *            P     - ESTIMATE OF FIRST SHAPE PARAMETER            *
C     *            Q     - ESTIMATE OF SECOND SHAPE PARAMETER           *
C     *            A     - ESTIMATE OF LOWER LIMIT (INITIAL VALUE INPUT)*
C     *            B     - ESTIMATE OF UPPER LIMIT (INITIAL VALUE INPUT)*
C     *            MEAN  - ARITHMATIC MEAN OF SAMPLE (INPUT)            *
C     *            SD    - STANDARD DEVIATION OF SAMPLE (INPUT)         *
C     *            T#    - SUM OF (X(I)**#),  #=1,..,4                  *
C     *            Y     - STANDARDIZED MEAN                            *
C     *            Z     - STANDARDIZED SAMPLE DEVIATION                *
C     *            ZSCNT - IMSL ROUTINE WHICH SOLVES SIMULTANEOUS       *
C     *                    NONLINEAR EQUATIONS BY THE SECANT METHOD     *
C     *            NPAR  - # PARAMETERS SOLVED FOR BY ZSCNT ( = # EQNS) *
C     *            NSIG  - # SIGNIFICANT DIGITS ZSCNT TO SOLVE FOR      *
C     *            ITMAX - MAXIMUM # ITERATIONS ZSCNT ALLOWED           *
C     *            PAR   - ARRAY OF PARAMETERS INPUTED BY ZSCNT         *
C     *            C     - ARRAY OF CONSTANTS INPUTED BY ZSCNT          *
C     *                    CONTAINS SAMPLE SKEWNESS, SAMPLE KURTOSIS    *
C     *            FCN   - SUBROUTINE CONTAINING EQUATIONS TO BE SOLVED *
C     *            FNORM - ZSCNT PARAMETER ( SEE IMSL MANUAL )          *
C     *            W     - ZSCNT WORKSPACE ( SEE IMSL MANUAL )          *
C     *            IER   - ZSCNT GENERATED ERROR INDICATOR              *
C     *                    ( SEE IMSL MANUAL )                          *
C     *                                                                 *
C     * I/O FILES: TAPE5 - INPUT, CONTAINS TRUE PARAMETERS AND RANDOM   *
C     *                    SAMPLES WITH ESTIMATED A + B, MEAN, STD DEV. *
C     *            TAPE6 - OUTPUT,CONTAINS TRUE PARAMETERS AND 4-MOMENT *
C     *                    PARAMETER ESTIMATES FOR EACH SAMPLE          *
C     *            TAPE7 - OUTPUT,CONTAINS TRUE PARAMETERS, AND 2-MOMENT*
C     *                    ESTIMATES OF P+Q, INTERPOLATED ESTIMATES OF  *
C     *                    A+B FOR EACH SAMPLE                          *
C     *            OUTPUT- CONTAINS MESSAGE ON # OF TIMES 4-MOMENT      *
C     *                    ESTIMATION FAILED                            *
```

```
C     *                                                                  *
C     * IMPORTANT: IMSL LIBRARY MUST BE ATTACHED BEFORE PROGRAM IS RUN  *
C     *              REVIEW IMSL MANUAL ON ZSCNT BEFORE RUNNING          *
C     *                                                                  *
C     ********************************************************************

      EXTERNAL ZSCNT,FCN
      DIMENSION X(50),PAR(2),C(2),W(42),F(2)
      REAL MEAN

C     READ IN TRUE PARAMETERS AND
C     WRITE THEM TO FILE
      READ(5,100) NREPS,N,PR,QR,AR,BR
  100 FORMAT(I4/I3/4(F10.6/))
      WRITE(6,106) NREPS
      WRITE(7,106) NREPS
      WRITE(6,101) N
      WRITE(7,101) N
  101 FORMAT(I3)
      WRITE(6,102) PR,QR,AR,BR
      WRITE(7,102) PR,QR,AR,BR
  102 FORMAT(4(F10.6,3X)/)
C     INITIALIZE DIVERGENCE COUNTER
      NDIV=0
C    ***** BEGIN LOOP FOR NREPS SAMPLES *****
      DO 999 J=1,NREPS
        READ(5,106) K
  106   FORMAT(I4)
C       INPUT SAMPLE POINTS
C       AND CALCULATE SUMS
        T2=0.0
        T3=0.0
        T4=0.0
        DO 1 I=1,N
          READ(5,103) X(I)
  103     FORMAT(F10.6)
          T2=T2+X(I)**2
          T3=T3+X(I)**3
          T4=T4+X(I)**4
    1   CONTINUE
C       INPUT CALCULATED VALUES
        READ(5,104) A,B,MEAN,SD
  104   FORMAT(F10.6/F10.6/F10.6/F10.6)
        T1=MEAN*N
C       FIND 2-MOMENT ESTIMATES OF P,Q
        Y=(MEAN-A)/(B-A)
        Z=SD/(B-A)
        P=(Y*Y*(1-Y))/(Z*Z) -Y
        Q=((1-Y)*(1-Y)*Y)/(Z*Z) -(1-Y)
C       WRITE SAMPLE INDEX, 2-MOMENT ESTIMATES OF
C       P+Q, AND INTERPOLATED ESTIMATES OF A+B TO FILE
        WRITE(7,106) K
        WRITE(7,102) P,Q
```

```
C       SET PARAMETERS FOR ZSCNT
        NPAR=2
        NSIG=3
        ITMAX=100
C       SOLVE 3RD AND 4TH MOMENT EQNS FOR P,Q
        PAR(1)=P
        PAR(2)=Q
C       CALCULATE SAMPLE SKEWNESS + KURTOSIS
C       USING EQNS 3.15 AND 3.16 IN THESIS
        C(1)=(N*N*T3-3*N*T2*T1+2*T1**3)/(SD**3*N*(N-1)*(N-2))
        C(2)=((N**3+N*N)*T4-4*(N*N+N)*T3*T1-3*(N*N-N)*T2**2
     +     +12*N*T2*T1**2-6*T1**4)/(SD**4*N*(N-1)*(N-2)*(N-3))
        CALL ZSCNT(FCN,NSIG,NPAR,ITMAX,C,PAR,FNORM,W,IER)
C       TEST FOR FEASIBILITY OF ESTIMATES
        IF( PAR(1).GT.0.0 .AND PAR(2).GT.0.0) THEN
C         IF ESTIMATES ARE FEASIBLE, SET EQUAL TO P+Q,
C         AND FIND A+B USING FIRST TWO MOMENTS
          P=PAR(1)
          Q=PAR(2)
          A=MEAN-SD*(P*(P+Q+1)/Q)**0.5
          B=(MEAN*(P+Q)-A*Q)/P
C         IF ESTIMATES OF A+B ARE INSIDE 1ST AND LAST
C         ORDER STATISTICS, USE ORDER STATISTICS AS ESTIMATES
          A=MIN(A,X(1))
          B=MAX(B,X(N))
        ELSE
C         IF ESTIMATES ARE INFEASIBLE, USE 2-MOMENT ESTIMATES
C         ( DON'T CHANGE VALUE OF P,Q) AND ADD 1 TO COUNTER
          NDIV=NDIV+1
        ENDIF
C       WRITE SAMPLE INDEX AND 4-MOMENT ESTIMATES
C       OF A,B,P,Q TO FILE
        WRITE(6,106) K
        WRITE(6,102) P,Q,A,B
C    ***** END LOOP *****
  999 CONTINUE
C     PRINT OUT # TIMES 4-MOMENT ESTIMATION FAILED
      PRINT*, ' NUMBER OF TIMES DID NOT CONVERGE =',NDIV
      STOP
      END
```

```
      SUBROUTINE FCN(PAR,F,NP,C)
C     ******************************************************************
C     *                                                                *
C     * PURPOSE:   EVALUATE EQUATIONS WHICH ZSCNT IS TRYING TO SOLVE    *
C     *                                                                *
C     * VARIABLES: PAR,C - SEE MAIN PROGRAM                            *
C     *            F     - ARRAY OF EQUATION VALUES AT THIS P,Q        *
C     *                    1ST EQN IS DIFFERENCE BETWEEN POP. + SAMPLE *
C     *                    SKEWNESS, 2ND EQN IS DIFFERENCE BETWEEN     *
C     *                    POP. + SAMPLE KURTOSIS                      *
C     *            NP    - NUMBER OF PARAMETERS, ALSO # OF EQUATIONS   *
C     *            PX    - SHORT NOTATION FOR PAR(1)                   *
C     *            QX    - SHORT NOTATION FOR PAR(2)                   *
C     *                                                                *
C     ******************************************************************

      DIMENSION PAR(NP),F(NP),C(2)

C     TEST FOR FEASIBILITY
      IF (PAR(1).LE.0.0 .OR. PAR(2).LE.0.0) THEN
C       SET  EQUATION VALUES TO ZERO
        F(1)=0.0
        F(2)=0.0
      ELSE
C       CHANGE TO SHORTER NOTATION
        PX=PAR(1)
        QX=PAR(2)
C       EVALUATE EQUATIONS
        F(1)=2.*(QX-PX)*((PX+QX+1)**0.5)/((PX+QX+2)*((PX*QX)**0.5))-C(1)
        F(2)=3.*(PX+QX+1)*(2*(PX+QX)**2+PX*QX*(PX+QX-6))
     +          /(PX*QX*(PX+QX+2)*(PX+QX+3))-C(2)
      ENDIF
      RETURN
      END
```

APPENDIX E

COMPUTER LISTING OF PROGRAM MDCVM

```
      PROGRAM MDCVM
C     ******************************************************************
C     *                                                                *
C     * WRITTEN BY 2LT DAVID E. BERTRAND AFIT/GOR-81D FOR MS THESIS    *
C     *                          DECEMBER 1981                         *
C     *                                                                *
C     * PURPOSE:    MINIMUM DISTANCE ESTIMATION OF THE FOUR PARAMETERS *
C     *             OF THE BETA DISTRIBUTION                           *
C     *                                                                *
C     * VARIABLES: NREPS  - # SAMPLES FOR WHICH ESTIMATION DONE (INPUT) *
C     *            N      - SAMPLE SIZE (INPUT)                        *
C     *            PR     - TRUE VALUE OF FIRST SHAPE PARAMETER (INPUT) *
C     *            QR     - TRUE VALUE OF SECOND SHAPE PARAMETER (INPUT)*
C     *            AR     - TRUE LOWER LIMIT OF DISTRIBUTION (INPUT)   *
C     *            BR     - TRUE UPPER LIMIT OF DISTRIBUTION (INPUT)   *
C     *            K      - SAMPLE INDEX (INPUT)                       *
C     *            X      - ARRAY OF SAMPLE POINTS (INPUT)             *
C     *            P      - ESTIMATE OF FIRST SHAPE PARAMETER          *
C     *            Q      - ESTIMATE OF SECOND SHAPE PARAMETER         *
C     *            A      - ESTIMATE OF LOWER LIMIT(INITIAL VALUE INPUT)*
C     *            B      - ESTIMATE OF UPPER LIMIT(INITIAL VALUE INPUT)*
C     *            MEAN   - ARITHMATIC MEAN OF SAMPLE (INPUT)          *
C     *            SD     - STANDARD DEVIATION OF SAMPLE (INPUT)       *
C     *            Y      - STANDARDIZED MEAN                          *
C     *            Z      - STANDARDIZED STANDARD DEVIATION            *
C     *            ZXMIN  - IMSL ROUTINE USED TO MINIMIZE DISTANCE     *
C     *            NPAR   - NUMBER OF VARIABLES INPUTED BY ZXMIN       *
C     *            NSIG   - # SIGNIFICANT DIGITS ZXMIN TO SOLVE FOR    *
C     *            MAXFN  - MAXIMUN # FUNCTIONAL EVALUATIONS BY ZXMIN  *
C     *            IOPT   - ZXMIN INPUT OPTION (SEE IMSL MANUAL)       *
C     *            PAR    - ARRAY OF PARAMETER VALUES USED BY ZXMIN    *
C     *            H,G,W  - ARRAYS USED BY ZXMIN (SEE IMSL MANUAL)     *
C     *            DISTPQ - SUBROUTINE TO FIND DISTANCE, P,Q INPUT     *
C     *            DISTAB - SUBROUTINE TO FIND DISTANCE, A,B INPUT     *
C     *            F      - DISTANCE VALUE: SEE SUBROUTINE             *
C     *            IER    - ZXMIN GENERATED ERROR MESSAGE              *
C     *                     ( SEE IMSL MANUAL )                        *
C     *                                                                *
C     * I/O FILES: TAPE5 - INPUT, CONTAINS TRUE PARAMETERS AND RANDOM  *
C     *                    SAMPLES WITH EST. A + B, MEAN, STD DEV.     *
C     *            TAPE6 - OUTPUT, CONTAINS TRUE PARAMETERS AND        *
C     *                    PARAMETER ESTIMATES FOR EACH SAMPLE         *
C     *                                                                *
C     * IMPORTANT: IMSL LIBRARY MUST BE ATTACHED BEFORE PROGRAM IS RUN *
C     *            REVIEW IMSL MANUAL ON ZXMIN AND MDBETA BEFORE RUNNING*
C     *            (MDBETA USED IN SUBROUTINE)                         *
C     *                                                                *
C     ******************************************************************
      COMMON P,Q,A,B,X(50),N
      EXTERNAL ZXMIN,MDBETA,DISTPQ,DISTAB
      DIMENSION PAR(2),H(3),G(2),W(6)
      REAL MEAN
```

```
C      INPUT TRUE PARAMETERS
       READ(5,100) NREPS,N,PR,QR,AR,BR
  100 FORMAT(I4/I3/4(F10.6/))
       WRITE(6,106) NREPS
       WRITE(6,101) N
  101 FORMAT(I3)
       WRITE(6,102) PR,QR,AR,BR
  102 FORMAT(4(F10.6)/)
  102 FORMAT(F10.6,3X)/)
C    ***** BEGIN LOOP FOR NREPS SAMPLES *****
       DO 99  J=1,NREPS
C        INPUT SAMPLE INDEX
         READ(5,106) K
  106    FORMAT(I4)
C        INPUT SAMPLE POINTS
         DO 1 I=1,N
           READ(5,103) X(I)
  103      FORMAT(F10.6)
    1      CONTINUE
C        INPUT CALCULATED VALUES
         READ(5,104) A,B,MEAN,SD
  104    FORMAT(F10.6/F10.6/F10.6/F10.6)
C        CALCULATE 2-MOMENT ESTIMATES OF P,Q
         Y=(MEAN-A)/(B-A)
         Z=SD/(B-A)
         P=(Y*Y*(1-Y))/(Z*Z)-Y
         Q=((1-Y)*(1-Y)*Y)/(Z*Z)-(1-Y)
C        SET ZXMIN PARAMETERS
         NPAR=2
         NSIG=3
         MAXFN=500
         IOPT=0
C        MINIMIZE DISTANCE FOR P,Q
         PAR(1)=P
         PAR(2)=Q
         CALL ZXMIN(DISTPQ,NPAR,NSIG,MAXFN,IOPT,PAR,H,G,F,W,IER)
         P=PAR(1)
         Q=PAR(2)
C        MINIMIZE DISTANCE FOR A,B
         PAR(1)=A
         PAR(2)=B
         CALL ZXMIN(DISTAB,NPAR,NSIG,MAXFN,IOPT,PAR,H,G,F,W,IER)
         A=PAR(1)
         B=PAR(2)
C        WRITE SAMPLE INDEX, ESTIMATES TO FILE
         WRITE(6,105) J
  105    FORMAT(I4)
         WRITE(6,102) P,Q,A,B
C    ***** END LOOP *****
  999 CONTINUE
       STOP
       END
```

```
      SUBROUTINE DISTPQ(NP,PAR,F)
C     ****************************************************************
C     *                                                              *
C     * PURPOSE:   FIND DISTANCE BETWEEN ESTIMATED CDF AND 1/N EDF   *
C     *            FOR VARIABLE P,Q KEEPING A,B FIXED                *
C     *                                                              *
C     * VARIABLES: NP     - NUMBER OF PARAMETERS: ALWAYS 2           *
C     *            PAR    - VECTOR OF PARAMETER VALUES P,Q           *
C     *            Y      - STANDARDIZED SAMPLE POINT                *
C     *            MDBETA - IMSL ROUTINE WHICH EVALUATES BETA CDF    *
C     *            Z      - VALUE OF BETA CDF AT POINT Y             *
C     *            IER    - MDBETA GENERATED ERROR MESSAGE           *
C     *                     ( SEE IMSL MANUAL)                       *
C     *            SUM    - DUMMY VARIABLE USED TO ADD UP DISTANCE   *
C     *            F      - DISTANCE VALUE AT THIS P,Q,A,B           *
C     *            P,Q,A,                                            *
C     *            B,X,N  - SEE MAIN PROGRAM                         *
C     *                                                              *
C     ****************************************************************

      COMMON P,Q,A,B,X(50),N
      INTEGER NP
      REAL PAR(NP),F,Y,Z,SUM

      SUM=0.0
      DO 92 I=1,N
C       STANDARDIZE SAMPLE POINT
        Y=(X(I)-A)/(B-A)
C       EVALUATE CDF
        CALL MDBETA(Y,PAR(1),PAR(2),Z,IER)
C       ADD TO SUM FOR DISTANCE
C        SEE EQN 3.25
        SUM=SUM+(Z-(2*I-1.)/(2.*N))**2
   92 CONTINUE
C     SET F EQUAL DISTANCE
      F=SUM
      RETURN
      END
```

```
      SUBROUTINE DISTAB(NP,PAR,F)
C     ******************************************************************
C     *                                                                *
C     * PURPOSE:    FIND DISTANCE BETWEEN ESTIMATED CDF AND 1/N EDF     *
C     *             FOR VARIABLE A,B KEEPING P,Q FIXED                 *
C     *                                                                *
C     * VARIABLES: NP     - NUMBER OF PARAMETERS: ALWAYS 2              *
C     *            PAR    - VECTOR OF PARAMETER VALUES A,B              *
C     *            Y      - STANDARDIZED SAMPLE POINT                   *
C     *            MDBETA - IMSL ROUTINE WHICH EVALUATES BETA CDF       *
C     *            Z      - VALUE OF BETA CDF AT POINT Y                *
C     *            IER    - MDBETA GENERATED ERROR MESSAGE              *
C     *                     ( SEE IMSL MANUAL)                          *
C     *            SUM    - DUMMY VARIABLE USED TO ADD UP DISTANCE      *
C     *            F      - DISTANCE VALUE AT THIS P,Q,A,B              *
C     *            P,Q,A,                                              *
C     *            B,X,N  - SEE MAIN PROGRAM                            *
C     *                                                                *
C     ******************************************************************

      COMMON P,Q,A,B,X(50),N
      INTEGER NP
      REAL PAR(NP),F,Y,Z,SUM

C     USE ORDER STATISTICS IF INTERPOLATED VALUES
C     ARE INSIDE 1ST AND LAST ORDER STATISTICS
      IF(PAR(1) .GT. X(1)) PAR(1)=X(1)
      IF(PAR(2) .LT. X(N)) PAR(2)=X(N)
      SUM=0.0
      DO 91 I=1,N
C       STANDARDIZE SAMPLE POINT
        Y=(X(I)-PAR(1))/(PAR(2)-PAR(1))
C       EVALUATE CDF
        CALL MDBETA(Y,P,Q,Z,IER)
C       ADD TO SUM FOR DISTANCE
C        SEE EQN 3.25
        SUM=SUM+(Z-(2*I-1.)/(2.*N))**2
   91 CONTINUE
C     SET F EQUAL DISTANCE
      F=SUM
      RETURN
      END
```

APPENDIX F

COMPUTER LISTING OF PROGRAM EVAL

```
      PROGRAM EVAL
C     ****************************************************************
C     *                                                              *
C     * WRITTEN BY 2LT DAVID E. BERTRAND AFIT/GOR-81D FOR MS THESIS   *
C     *                        DECEMBER 1981                         *
C     *                                                              *
C     * PURPOSE:   EVALUATE A SET OF ESTIMATES ON THE BETA DISTRIBUTION *
C     *              - CALCULATE THE MEAN SQUARE ERROR OF THE SET FOR  *
C     *                EACH OF THE FOUR PARAMETERS                   *
C     *              - FIND THE CVM DISTANCE BETWEEN THE ESTIMATED AND *
C     *                TRUE CDF, AND FIND MEAN AND STD DEV OF CVM FOR SET*
C     *                                                              *
C     * VARIABLES: NREPS  - # OF ESTIMATES OF EACH PARAMETER IN SET   *
C     *                     (INPUT)                                  *
C     *            N      - SIZE OF SAMPLES ON WHICH ESTIMATES ARE BASED*
C     *                     (INPUT)                                  *
C     *            PR     - TRUE VALUE OF FIRST SHAPE PARAMETER (INPUT) *
C     *            QR     - TRUE VALUE OF SECOND SHAPE PARAMETER (INPUT)*
C     *            AR     - TRUE VALUE OF LOWER LIMIT (INPUT)         *
C     *            BR     - TRUE VALUE OF UPPER LIMIT (INPUT)         *
C     *            X      - ARRAY CONTAINING GAUSSIAN QUADRATURE POINTS *
C     *                     (INPUT)                                  *
C     *            W      - ARRAY CONTAINING GAUSSIAN QUADRATURE WEIGHTS*
C     *                     (INPUT)                                  *
C     *            SUM1   - DUMMY VAR. USED TO SUM CVM STATS OF EACH  *
C     *                     REPITITION                               *
C     *            SEP    - SQUARED ERROR OF P IN THIS REPITITION     *
C     *            SEQ    - SQUARED ERROR OF Q IN THIS REPITITION     *
C     *            SEA    - SQUARED ERROR OF A IN THIS REPITITION     *
C     *            SEB    - SQUARED ERROR OF B IN THIS REPITITION     *
C     *            P      - ESTIMATE OF FIRST SHAPE PARAMETER (INPUT) *
C     *            Q      - ESTIMATE OF SECOND SHAPE PARAMETER (INPUT) *
C     *            A      - ESTIMATE OF LOWER LIMIT (INPUT)           *
C     *            B      - ESTIMATE OF UPPER LIMIT (INPUT)           *
C     *            SUM    - DUMMY VAR FOR EVAL OF INTEGRAL BY QUADRATURE*
C     *            Y      - STANDARDIZED QUADRATURE POINT             *
C     *            MDBETA - IMSL ROUTINE WHICH EVALUATES STD BETA CDF *
C     *            IER    - MDBETA GENERATED ERROR INDICATOR          *
C     *                     ( SEE IMSL MANUAL )                      *
C     *            FN     - VALUE OF ESTIMATED CDF AT QUADRATURE POINT *
C     *            FR     - VALUE OF TRUE CDF AT QUADRATURE POINT     *
C     *            BETA   - BETA FUNCTION - SEE EQN 3.??              *
C     *            GAMMA  - IMSL ROUTINE WHICH EVALUATES THE GAMMA FCN *
C     *            F      - VALUE OF TRUE PDF AT QUADRATURE POINT     *
C     *            CVM    - ARRAY CONTAINING CVM DISTANCE BETWEEN     *
C     *                     ESTIMATED AND TRUE CDF FOR EACH REPITITION *
C     *            MSE    - MEAN SQUARE ERROR OF P                    *
C     *            MSEQ   - MEAN SQUARE ERROR OF Q                    *
C     *            MSEA   - MEAN SQUARE ERROR OF A                    *
C     *            MSEB   - MEAN SQUARE ERROR OF B                    *
C     *            MCVM   - MEAN OF THE CVM DISTANCES                 *
C     *            SDCVM  - STD DEV OF THE CVM DISTANCES              *
```

```
C     *                                                                   *
C     * I/O FILES: TAPE6  - INPUT, CONTAINS TRUE PARAMETER VALUES AND     *
C     *                     ESTIMATES FOR EACH REPITITION                 *
C     *            TAPE7  - OUTPUT, CONTAINS MSE'S AND MEAN + STD DEV     *
C     *                     OF CVM DISTANCES                              *
C     *            INPUT  - CONTAINS 8 POSITIVE QUADRATURE POINTS AND     *
C     *                     WEIGHTS FOR 16 POINT GAUSSIAN QUADRATURE      *
C     *                                                                   *
C     * IMPORTANT: IMSL LIBRARY MUST BE ATTACHED BEFORE PROGRAM IS RUN    *
C     *           REVIEW IMSL MANUAL ON MDBETA AND GAMMA BEFORE RUNNING*
C     *                                                                   *
C     *********************************************************************

      EXTERNAL MDBETA,GAMMA
      REAL X(8,2),W(8),MCVM,MSEP,MSEQ,MSEA,MSEB,CVM(1000)

C     INPUT TRUE PARAMETER VALUES
      READ(6,100) NREPS,,N,PR,QR,AR,BR
  100 FORMAT(I4/I3/4(F10.6.3X)/)
C     INPUT QUADRATURE POINTS AND WEIGHTS
      DO 1 I=1,8
        READ*, X(I,1),W(I)
        X(I,2)=-1.*X(I,1)
C       TRANSLATE QUAD PTS TO (AR,BR) INTERVAL
        DO 2 J=1,2
          X(I,J)=((BR-AR)/2)*X(I,J)+(BR+AR)/2
    2   CONTINUE
    1 CONTINUE
C     INITIALIZE SUMS
      SUM1=0.0
      SEP=0.0
      SEQ=0.0
      SEA=0.0
      SEB=0.0
C    ***** BEGIN LOOP FOR NREPS REPITITIONS *****
      DO 999 K=1,NREPS
C       INPUT PARAMETER ESTIMATES
        READ(6,101) P,Q,A,B
  101    FORMAT(/4(F10.6,3X)/)
C       EVALUATE CVM INTEGRAL BY QUADRATURE
        SUM=0.0
        DO 888 J=1,2
          DO 777 I=1,8
C           STANDARDIZE QUADRATURE POINT
C           USING ESTIMATED VALUES OF A + B
            Y=(X(I,J)-A)/(B-A)
C           RESET STANDARDIXED QUAD PT IF
C           IT IS OUTSIDE ESTIMATED RANGE
            IF(Y.LT.0.0) Y=0.0
            IF(Y.GT.1.0) Y=1.0
C           EVALUATE EST. BETA CDF
            CALL MDBETA(Y,P,Q,FN,IER)
C           STANDARDIZE QUADRATURE POINT
C           USING TRUE VALUES OF A + B
```

```
              Y=(X(I,J)-AR)/(BR-AR)
C             EVALUATE TRUE BETA CDF
              CALL MDBETA(Y,PR,QR,FR,IER)
C             EVALUATE TRUE BETA PDF
              BETA=GAMMA(PR)*GAMMA(QR)/GAMMA(PR+QR)
              F=(1/BETA)*(X(I,J)-AR)**(PR-1)*(BR-X(I,J))**(QR-1)
     *          /(BR-AR)**(PR+QR-1)
C             ADD TO SUM FOR EVAL. OF INTEGRAL
              SUM=SUM+W(I)*(FN-FR)**2*F
  777       CONTINUE
  888     CONTINUE
C         CALCULATE CVM STATISTIC
          CVM(K)=N*((BR-AR)/2)*SUM
C         ADD TO SUMS FOR CVM, SQUARED ERRORS
          SUM1=SUM1+CVM(K)
          SEP =SEP+(PR-P)**2
          SEQ =SEQ+(QR-Q)**2
          SEA =SEA+(AR-A)**2
          SEB =SEB+(BR-B)**2
C       ***** END LOOP *****
  999 CONTINUE
C     CALCULATE MEAN SQUARE ERRORS, MEAN CVM
      MSEP= SEP/NREPS
      MSEQ= SEQ/NREPS
      MSEA= SEA/NREPS
      MSEB= SEB/NREPS
      MCVM=SUM1/NREPS
C     CALCULATE STD DEV OF CVM STATISTICS
      SUM=0.0
      DO 3 K=1,NREPS
        SUM=SUM+(CVM(K)-MCVM)**2
    3 CONTINUE
      SDCVM=(SUM/NREPS)**0.5
C     WRITE RESULTS TO FILE
      WRITE(7,103) MSEP,MSEQ,MSEA,MSEB,MCVM,SDCVM
  103 FORMAT(4(F10.6,3X)/2(F10.6,3X))
      STOP
      END
```

Vita

David E. Bertrand was born in Lowell, Massachusetts on 24 September 1958. He graduated from Billerica Memorial High School in Billerica, Massachusetts, in June of 1976. He then enrolled in the University of Lowell, pursuing a degree in Mathematics. In May, 1980, he received his B. S. in Mathematics and also was commissioned a Second Lieutenant in the United States Air Force through the ROTC program. He was assigned directly to the School of Engineering, Air Force Institute of Technology, in June of 1980.

David Bertrand will be married to Sharon Ann Sorano in their hometown of Bellerica on 3 April 1982.

Permanent Address: 353 Concord Rd.
Billerica, MA 01821

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFIT/GOR/81D-1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) COMPARISON OF ESTIMATION TECHNIQUES FOR THE FOUR PARAMETER BETA DISTRIBUTION		5. TYPE OF REPORT & PERIOD COVERED MS Thesis
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) David E. Bertrand 2Lt. USAF		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Air Force Institute of Technology (AFIT/EN) Wright-Patterson AFB, Ohio 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N/A
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE December 1981
		13. NUMBER OF PAGES 77
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

15 APR 1982

18. SUPPLEMENTARY NOTES

Approved for public release, AFR 190-17

~~Fredrick Lynch, Major, USAF~~ [signature]

~~Director of Public Affairs~~

Dean for Research and Professional Development

Air Force Institute of Technology (ATC)

Wright-Patterson AFB, OH 45433

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Beta Distribution

Estimation

Minimum Distance Estimation

Cramer - von Mises Distance

Monte Carlo Analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This thesis compares three estimation techniques in application to the beta distribution: method of moments, maximum likelihood, and minimum distance. The four parameter version of the beta distribution is used; it has two shape parameters, and upper and lower limit parameters. Linear interpolation on order statistics is used to find initial estimates of the limits. The classical

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 continued:

estimation procedures, method of moments and maximum likelihood, are applied through procedures found in the literature. A newer technique, minimum distance, is applied for the first time to the beta distribution.

Comparison of estimation techniques is accomplished using Monte Carlo analysis. Five sample sizes are considered — 4, 8, 12, 16, and 20 — and three pairs of shape parameters — (3,3) , (9,4) , and (1,2) — for a total of fifteen cases. One thousand samples are generated for each case, and each estimation technique is then applied to all samples. Two effectiveness measures are used; they are the mean square error of each parameter estimate, and the Cramer - von Mises distance between the estimated and the true distribution. These effectiveness measures are compared in each case to determine which technique provides the best overall effectiveness..

END

DATE
FILMED

7 82

DTIC